ICE-B 2008

INTERNATIONAL CONFERENCE ON E-BUSINESS

Proceedings

PORTO, PORTUGAL · JULY 26-29, 2008
BRIEF CONTENTS

INVITED SPEAKERS ........................................................................................................................ IV
TUTORIAL ........................................................................................................................................ IV
ORGANIZING AND STEERING COMMITTEES ................................................................................ V
PROGRAM COMMITTEE ............................................................................................................... VI
 AUXILIARY REVIEWERS ......................................................................................................... VIII
SELECTED PAPERS BOOK ........................................................................................................ VIII
FOREWORD ............................................................................................................................. IX
CONTENTS ............................................................................................................................... XI
INVITED SPEAKERS

David A. Marca
University of Phoenix
U.S.A.

Yaakov Kogan
AT&T Labs
U.S.A.

Hsiao-Hwa Chen
National Sun Yat-Sen University
Taiwan

Nuno Borges Carvalho
Instituto de Telecomunicações / Universidade de Aveiro
Portugal

Ueli Maurer
ISwiss Federal Institute of Technology (ETH)
Switzerland

Bart Preneel
University of Leuven
Belgium

Ingemar Cox
University College London
U.K.

TUTORIAL

PATENTABILITY OF E-BUSINESS AND COMPUTER-IMPLEMENTED INVENTIONS AT THE EPO

Falk Giemsa, European Patent Office, Germany
ORGANIZING AND STEERING COMMITTEES

CONFERENCE CO-chairs
Joaquim Filipe, Polytechnic Institute of Setúbal / INSTICC, Portugal
Mohammad S. Obaidat, Monmouth University, U.S.A.

PROGRAM CO-chairs
Marten van Sinderen, University of Twente, The Netherlands
Boris Shishkov, University of Twente, The Netherlands
David A. Marca, University of Phoenix, U.S.A.

PROCEEDINGS PRODUCTION
Helder Coelhas, INSTICC, Portugal
Vera Coelho, INSTICC, Portugal
Andreia Costa, INSTICC, Portugal
Bruno Encarnação, INSTICC, Portugal
Bárbara Lima, INSTICC, Portugal
Vitor Pedrosa, INSTICC, Portugal
Vera Rosário, INSTICC, Portugal

CD-ROM PRODUCTION
Paulo Brito, INSTICC, Portugal

GRAPHICS PRODUCTION
Helder Coelhas, INSTICC, Portugal

SECRETARIAT, WEBDESIGNER AND WEBMASTER
Mónica Saramago, INSTICC, Portugal
PROGRAM COMMITTEE

Geetha Abeyesinghe, Middlesex University, U.K.
Ajith Abraham, Norwegian University of Science and Technology, Norway
Fahim Akhter, Zayed University, U.A.E.
Damminda Alahakoon, Monash University, Australia
Antonia Albani, Delft University of Technology, The Netherlands
Sarmad Alshawi, Brunel University, U.K.
Panagiotes Anastasiades, University of Crete, Greece
José Luis Arjona, University of Huelva, Spain
Anteneh Ayanso, Brock University, Canada
Gilbert Babin, HEC Montréal, Canada
Ladjel Bellatreche, Poitiers University, France
Morad Benyoucef, University of Ottawa, Canada
Hans Björnsson, Chalmers University of Technology, Sweden
Peter Bodorik, Dalhousie University, Canada
Indranil Bose, University of Hong Kong, Hong Kong
Vesna Bosilj-Vuksic, University of Zagreb, Faculty of Economics and Business, Croatia
Christos Bouras, University of Patras, Greece
Stephane Bressan, National University of Singapore, Singapore
Rongzeng Cao, IBM China Research Lab, China
Barbara Carminati, University of Insubria, Italy
Teuta Cata, Northern Kentucky University, U.S.A.
Michelangelo Ceci, University of Bari, Italy
Wojciech Cellary, Poznan University of Economics, Poland
Patrick Y.K. Chau, The University of Hong Kong, Hong Kong
Michael Chau, The University of Hong Kong, Hong Kong
Harry Chen, Image Matters LLC/University of Maryland, Baltimore County, U.S.A.
Dickson Chiu, Dickson Computer Systems, Hong Kong
Soon Ae Chun, City University of New York, U.S.A.
Jen-Yao Chung, IBM T. J. Watson Research Center, U.S.A.
Oscar Corcho, Universidad Politécnica de Madrid, Spain
Alfredo Cuzzocrea, University of Calabria, Italy
George Dafoulas, Middlesex University, U.K.
Hepu Deng, RMIT University, Australia
Claudia Diaz, K. U. Leuven, Belgium
Asuman Dogac, Middle East Technical University, Turkey
Schahram Dustdar, T. U. Wien, Austria
Joerg Evermann, Memorial University of Canada, Canada
Jinan Fiaidhi, Lakehead University, Canada
Xiang Fu, Georgia Southwestern State University, U.S.A.
George Giaglis, Athens University of Economics and Business, Greece
Paul Grefen, Eindhoven University of Technology, The Netherlands
Volker Gruhn, University of Leipzig, Germany
Haresh Gurnani, University of Miami, U.S.A.
Mohand-Said Hadid, University Lyon 1, France
Milena Head, McMaster University, Canada
Vlatka Hlupic, University of Westminster, U.K.
Birgit Hofreiter, University of Technology Sydney, Australia
Andreas Holzinger, IMI, Research Unit HCI, Medical University Graz, Austria
Christian Huemer, Vienna University of Technology, Austria
Patrick C. K. Hung, University of Ontario Institute of Technology, Canada
Takayuki Ito, Nagoya Institute of Technology, Japan
Arun Iyengar, IBM T.J. Watson Research Center, U.S.A.
Nallani Iyengar, Vellore Institute of Technology University, India
James Joshi, University of Pittsburgh, U.S.A.
Matjaz B. Juric, University of Maribor, Slovenia
Sherif Kamel, The American University in Cairo, Egypt
Kay Hooi Keoy, Bradford University School of Management, U.K.
David Kreps, University of Salford, U.K.
Anton Lavrin, The Technical University of Kosice, Slovakia
Lundy Lewis, Southern New Hampshire University, U.S.A.
Dahui Li, University of Minnesota Duluth, U.S.A.
Yinseng Li, Fudan University, China
Chin Lin, South China University of Technology, China
Tokuro Matsuo, Yamagata University, Japan
Jan Mendling, QUT Brisbane, Australia
Brian Mennecke, Iowa State University, U.S.A.
Adrian Mocan, Digital Enterprise Research Institute, Austria
Sabah Mohammed, Lakehead University, Canada
Ali Reza Montazemi, McMaster University, Canada
Wee-Keong Ng, Nanyang Technological University, Singapore
Dan O'Leary, University of Southern California, U.S.A.
Georgios Papamichail, National Centre of Public Administration and Local Government, Greece
Cesare Pautasso, University of Lugano, Switzerland
Krassie Petrova, Auckland University of Technology, New Zealand
Pascal Poncelet, LGI2P/EMA, France
Pak-Lok Poon, The Hong Kong Polytechnic University, Hong Kong
Philippos Pouyioutas, University of Nicosia, Cyprus
Dimitris Rigas, University of Bradford, U.K.
David Ruiz, University of Seville, Spain
Jarogniew Rykowski, Poznan University of Economics, Poland
Demetrios Sampson, University of Piraeus, Greece
Hossein Sharifi, Liverpool University Management School, U.K.
Quan Z. Sheng, The University of Adelaide, Australia
Mario Spremic, University of Zagreb, Croatia
Katarina Stanoevska-Slabeva, University of St. Gallen, Switzerland
York Sure, SAP Research, CEC Karlsruhe, Germany
Paula Swatman, University of South Australia, Australia
Ramayah T., Universiti Sains Malaysia, Malaysia
Thompson Teo, National University of Singapore, Singapore
Thanassis Tiropoulos, University of Southampton, U.K.
David Trastour, HP Labs, U.K.
Roland Traunmüller, University Linz, Austria
Jan Vanthienen, Katholieke Universiteit Leuven, Belgium
Tomas Vitvar, National University of Ireland, Ireland
Adam Vrechopoulos, Athens University of Economics and Business, Greece
Yan Wang, Macquarie University, Australia
Krzysztof Wecel, Poznan University of Economics, Poland
Michael Weiss, Carleton University, Canada
Erik Wilde, UC Berkeley, U.S.A.
Jongwook Woo, California State University Los Angeles, U.S.A.
Lai Xu, SAP Research, Switzerland
Benjamin Yen, The University of Hong Kong, Hong Kong
Soe-Tsyr Yuan, National Chengchi University, Taiwan
Guangquan Zhang, University of Technology Sydney, Australia
A number of selected papers presented at ICETE 2008 will be published by Springer-Verlag in a CCIS Series book. This selection will be done by the Conference Co-chairs and Program Co-chairs, among the papers actually presented at the conference, based on a rigorous review by the ICETE 2008 program committee members.
We warmly welcome you to ICE-B 2008 - the International Conference on E-Business, which is held, this year, in Porto, Portugal. This conference reflects a growing effort to increase the dissemination of recent research results among professionals who work on the e-business field. ICE-B is integrated as one of the modules of the ICETE conference, which occurs concurrently with ICE-B.

The major goal of ICETE is to bring together researchers, engineers and practitioners interested in information and communication technologies, including e-business, wireless networks and information systems, security and cryptography, signal processing and multimedia applications. These are the main knowledge areas that define the four component conferences, namely: ICE-B, SECRYPT, SIGMAP and WINSYS, which together form the ICETE joint conference.

In the program for ICETE, we have included keynote lectures, tutorials, papers, and posters to present the widest possible view on these technical areas. With its four tracks, we expect to appeal to a global audience of the engineers, scientists, business practitioners and policy experts, interested in the research topics of ICETE. All tracks focus on research related to real world applications and rely on contributions not only from Academia, but also from industry, with different solutions for end-user applications and enabling technologies, in a diversity of communication environments. The four volume set of proceedings demonstrate a number of new and innovative solutions for e-business and telecommunication, and demonstrate the vitality of these research areas.

ICETE has received 440 papers in total, with contributions from more than 40 different countries, from all continents, which demonstrates the success and global dimension of ICETE 2008. To evaluate each submission, a double blind paper evaluation method was used: each paper was reviewed by at least two experts from the International Program Committee, in a double-blind review process, and most papers had 3 reviews or more. In the end, 174 papers were selected for oral presentation and publication, corresponding to a 39% acceptance ratio. Of these only 77 were accepted as full papers (17% of submissions) and 97 as short papers. Additionally, 87 papers were accepted for poster presentation. These acceptance ratios demonstrate that ICETE 2008 strives to achieve a high quality standard which we will keep and enhance in order to ensure the success of next year conference, to be held in Milan/Italy. A short list of about thirty papers will be also selected to appear in a book that will be published by Springer.

We would like to emphasize that ICETE 2008 includes several outstanding keynote lectures, which are relevant to today’s lines of research and technical innovation. These talks are presented by distinguished researchers who are internationally renowned experts in all ICETE areas, and their contributions heighten the overall quality of our Conference.

A successful conference involves more than paper presentations; it is also a meeting place, where ideas about new research projects and other ventures are discussed and debated. Therefore, a social event including a conference diner/banquet has been planned for the evening of July 28 in order to promote this kind of social networking.
We would like to express our thanks, first of all, to all authors including those whose papers were not included in the program. Next, we would like to thank all the members of the program committee and reviewers, who helped us with their expertise, dedication and time. We would also like to thank the invited speakers for their invaluable contribution, in sharing their vision and knowledge. Lastly, but certainly not least, we give our deep appreciation to the secretariat and to all the other members of the organizing committee, whose diligence in dealing with all organizational issues were essential to a collaborative effort of a dedicated and highly capable team.

We hope that you will find these proceedings interesting and to be a helpful reference in the future for all those who need to address the areas of e-business and telecommunications.

Enjoy the program and your stay in Porto.

**Marten van Sinderen**  
University of Twente, The Netherlands

**Boris Shishkov**  
University of Twente, The Netherlands

**David A. Marca**  
University of Phoenix, U.S.A.
## CONTENTS

### INVITED SPEAKERS

#### KEYNOTE LECTURES

<table>
<thead>
<tr>
<th>Title</th>
<th>Speaker</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>E-BUSINESS INNOVATION - Surviving the Coming Decades</strong></td>
<td>David A. Marca</td>
<td>IS-5</td>
</tr>
<tr>
<td><strong>IMPROVING RELIABILITY IN COMMERCIAL IP NETWORKS</strong></td>
<td>Yaakov Kogan</td>
<td>IS-17</td>
</tr>
<tr>
<td><strong>CRYPTOGRAPHIC ALGORITHMS - Successes, Failures and Challenges</strong></td>
<td>Bart Preneel</td>
<td>IS-21</td>
</tr>
<tr>
<td><strong>WATERMARKING, STEGANOGRAPHY AND CONTENT FORENSICS</strong></td>
<td>Ingemar J. Cox</td>
<td>IS-29</td>
</tr>
<tr>
<td><strong>RETHINKING DIGITAL SIGNATURES</strong></td>
<td>Ueli Maurer</td>
<td>IS-31</td>
</tr>
<tr>
<td><strong>THE IMPORTANCE OF METROLOGY IN WIRELESS COMMUNICATION SYSTEMS - From AM/FM to SDR Systems</strong></td>
<td>Nuno Borges Carvalho</td>
<td>IS-35</td>
</tr>
<tr>
<td><strong>NEXT GENERATION CDMA TECHNOLOGIES FOR FUTURISTIC WIRELESS COMMUNICATIONS</strong></td>
<td>Hsiao-Hwa Chen</td>
<td>IS-37</td>
</tr>
</tbody>
</table>

#### TUTORIAL

<table>
<thead>
<tr>
<th>Title</th>
<th>Speaker</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>PATENTABILITY OF E-BUSINESS AND COMPUTER-IMPLEMENTED INVENTIONS AT THE EPO</strong></td>
<td>Falk Giemsa</td>
<td>IS-41</td>
</tr>
</tbody>
</table>

### COMMUNICATION AND SOFTWARE TECHNOLOGIES AND ARCHITECTURES

#### FULL PAPERS

<table>
<thead>
<tr>
<th>Title</th>
<th>Authors</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>A PERFORMANCE EVALUATION OF AN ULTRA-THIN CLIENT SYSTEM</strong></td>
<td>Colin Pattinson and Tabir Siddiqui</td>
<td>5</td>
</tr>
<tr>
<td><strong>USING CO-OCCURRENCE TO CLASSIFY UNSTRUCTURED DATA IN TELECOMMUNICATION SERVICES</strong></td>
<td>Motoi Iwashita, Ken Nishimatsu and Shinsuke Shinogawa</td>
<td>12</td>
</tr>
</tbody>
</table>

#### SHORT PAPERS

<table>
<thead>
<tr>
<th>Title</th>
<th>Authors</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>FLEXIBLE DATA SEARCHS USING CONDITION FORMULAS</strong></td>
<td>Toshio Kodama, Toshiyasu L., Kuniti and Yoichi Seki</td>
<td>21</td>
</tr>
</tbody>
</table>
ANALYZING DECENTRALIZED GOVERNABILITY OF BUSINESS PROCESSES BY EXTENDED PETRI NETS AND MODAL LOGICS
Takashi Hattori, Hiroshi Kawakami, Osamu Katai and Takayuki Shinse

AN ARCHITECTURE FOR DYNAMIC INVARIANT GENERATION IN WS-BPEL WEB SERVICE COMPOSITIONS
M. Palomo Duarte, A. García Dominguez and I. Medina Bulo

A NEW REINFORCEMENT SCHEME FOR STOCHASTIC LEARNING AUTOMATA - Application to Automatic Control
Florin Stoica, Emil M. Popa and Iulian Pah

BUSINESS INTELLIGENCE THROUGH REAL-TIME TRACKING - Using a Location System Towards Behaviour Pattern Extraction
Pedro Abreu, Vasco Vinhas and Pedro Mendes

A FRAMEWORK FOR DYNAMIC KNOWLEDGE ACQUISITION
Ana Aguilera and Alberto Subero

PRES – PERSONALIZED EVALUATION SYSTEM IN A WEB COMMUNITY - A Conceptual Model Designed to Evaluate Reputation in Order to Achieve a Personalised View on the System for Each User
Lenuta Alboaie

POSTERS

A STUDY OF THE EFFECTIVENESS OF “WAKE UP ON LAN” AS A MEANS OF POWER MANAGEMENT
Colin Pattinson and Linton Robinson

ALGORITHM AND AN ELEVATOR CONTROL SYSTEM EXAMPLE FOR CTL MODEL UPDATE
Cacovean Laura Florentina, Pah Iulian, Popa Emil Marin and Brumar Cristina Ioana

COMMON TEXTILE VOCABULARIES AND DOCUMENTS - A Conceptual Foundation of a Globally Interoperable Textile e-Marketplace
Jingzhi Guo and Zhuo Hu

PROPOSAL OF AN ARCHITECTURE FOR DIGITAL CITIES CREATION - Proposal of an Architecture P2P for Digital Cities Creation
André M. Panhan, Denny G. Santos and Leonardo S. Mendes

OPEN SOURCE SOFTWARE AND LEVERAGING OF BUSINESS EFFECTIVENESS IN SMES - A Case Study
Steven Butler, Dotun Adebanjo and Hassan Ismail

AN ONTOLOGY-BASED ARCHITECTURE FOR MULTI-AGENT SYSTEM ENVIRONMENT
Roberto Paitano, Anna Lisa Guido and Enrico Pulimeno

SECURE IT/TELCO ENVIRONMENT PLANNING MADE EASY - A Concept of a Tool for Planning Secure IT/Telco Infrastructure and Applications
Wolfgang Haidegger

WEB AND MOBILE BUSINESS SYSTEMS AND SERVICES

FULL PAPERS

MOBILE TOURISM SERVICES - Experiences from Three Services on Trial
Niklas Eriksson and Peter Strandvik
ADOPTION VERSUS USE DIFFUSION - Predicting User Acceptance of Mobile TV in Flanders
Tom Evens, Lieven De Marez and Dimitri Schuurman

ADOPTION OF MOBILE SERVICES IN FINLAND - Conceptual Model and Application-based Case Study
Mari Ervasti and Heli Helaakoski

SHORT PAPERS

THE ROLE OF AUDIO-VISUAL METAPHORS IN AIDING THE COMMUNICATION OF CUSTOMER KNOWLEDGE - User Satisfaction Perspective
Dimitrios I. Rigas and Mutlaq B. Alotaibi

A PERVASIVE NUTRITIONAL MONITORING AND ADVISE SYSTEM - NutriMe
Vitor Basto Fernandes, João Varajão and António Cunha

REPLICATION OF WEB SERVICES FOR QOS GUARANTEES IN WEB SERVICE COMPOSITION
Dirk Thissen and Thomas Brambring

AN E-VORTAL FOR THE PORTUGUESE BAKING INDUSTRY - Requirements Model
João Varajão, Jorge Gouveia and Paula Oliveira

ON EXPLORING CONSUMERS’ TECHNOLOGY FORESIGHT CAPABILITIES - An Analysis of 4,000 Mobile Service Ideas
Petteri Alahuhta, Pekka Abrahamsson and Antti Nummiaho

AN ANALYSIS OF CONTEXT-AWARENESS IN COMMERCIAL MOBILE SERVICES
Ana M. Bernardos, Daniel Marcos and José R. Casar

AN ACCESS-CONTROL MODEL FOR MOBILE COMPUTING WITH SPATIAL CONSTRAINTS - Location-aware Role-based Access Control with a Method for Consistency Checks
Michael Decker

A PLATFORM FOR INVESTIGATING EFFECTIVENESS FOR STATIC, ADAPTABLE, ADAPTIVE, AND MIXED-INITIATIVE ENVIRONMENTS IN E-COMMERCE
Khaled Al-Omar and Dimitris Rigas

AN ENHANCED SERVICE PROVIDER COMMUNICATION INTERFACE WITH CLIENT PRIORITIZATION - Case Study on Fast-food Chain Restaurants
Slobodan Lukovic, Nikola Puzovic and Milos Stanićević

MOBILE BUSINESS EXPERT ADVISOR
Danco Davcev, Marjan Arsic and Dalibor Ilievski

A MOBILE BUSINESS PROCESS DEPLOYMENT FRAMEWORK FOR DEVICE INDEPENDENCE AND CONTEXT-AWARE ENVIRONMENTS
Torab Torabi, Saqib Ali and Hassan Ali

GIS-BASED MAP GENERATION USING NEW SURVEY TECHNIQUES
Balqies Sadoun and Omar Al-Bayari

POSTERS

eCT: THE B2C E-COMMERCE TOOLKIT FOR THE WEBCOMFORT PLATFORM
Frederico de Carvalho Baptista, João de Sousa Saraiva and Alberto Rodrigues da Silva
SURVEY OF CONSUMERS’ DECISION MAKING PROCESS FOR ONLINE MUSIC SERVICE
Ayako Hiramatsu, Takahiro Yamasaki and Kazuo Nose 229

TYPICAL PROBLEMS WITH DEVELOPING MOBILE APPLICATIONS FOR HEALTH CARE - Some Lessons Learned from Developing User-centered Mobile Applications in a Hospital Environment
Andreas Holzinger, Martin Hüller, Marcus Blosé and Berndt Urlesberger 235

A STUDY OF FACTOR AFFECTING CUSTOMER SWITCHING BEHAVIOR OF MOBILE TELECOMMUNICATION 3.5G SERVICES
Jungwoo Lee, Eok Bae and Su Jung Sung 241

ADOPTION OF NEAR FIELD COMMUNICATION TECHNOLOGY IN BUSINESS TO CONSUMER SERVICES
Arto Wallin, Juha Häikö and Jaana Määttä 247

TAXONOMY FOR MOBILE TERMINALS - A Selective Classification Scheme
Gunther Schiefer and Michael Decker 255

IMPLEMENTING TRADING AGENTS FOR ADAPTABLE AND EVOLUTIVE UI-COTS COMPONENTS ARCHITECTURES
José Andrés Asensio, Luis Iribarne, Nicolás Padilla and Rosa Ayala 259

BUSINESS AND SOCIAL APPLICATIONS

FULL PAPERS

DEALING WITH BUSINESS PROCESS EVOLUTION USING VERSIONS
Mohamed Amine Chaâbane, Eric Andonoff, Lutfi Bouzguenda and Rafik Bouaziz 267

A THEORY-DRIVEN FRAMEWORK FOR CONSUMERS TO ADOPT M-COMMERCE DEVICES
Vincent Cho, Humphry Hung and Y. H. Wong 279

IT APPLICATIONS IN PRODUCTION PLANNING AND CONTROL - A Survey of Medium Sized Business in German-speaking Europe
Jakob Lewandowski, Matthias Buhl and Burkhard Kittl 285

CONDITIONS FOR TECHNOLOGY ACCEPTANCE - Broadening the Scope of Determinants of ICT Appropriation
Pieter Verdegem and Lieven De Marez 292

BARRIERS TO MOBILE BANKING ADOPTION - A Cross-national Study
Tommi Laukkanen and Pedro Cruz 300

BUSINESS AND TECHNICAL WORKFLOWS FOR E-BUSINESS IN A VIRTUAL CLUSTER OF ISPS
Jane Hall and Klaus-Peter Eckert 307

FACTORS AFFECTING THE USAGE OF T-GOVERNMENT SERVICES - An Exploratory Study
Michele Cornacchia, Filomena Papa, Stefano Livi, Bartolomeo Sapio, Enrico Nicolò and Gaetano Bruno 315

BUSINESS PROCESSES MANAGEMENT USING PROCESS ALGEBRA AND RELATIONAL DATABASE MODEL
Kelly Rosa Braghetto, João Eduardo Ferreira and Calvin Pu 323

ON THE USE OF “QUALIFIED” DIGITAL SIGNATURES
Maurizio Adriano Strangio 334
RESEARCHING SEARCH - A Study into Search Engine Marketing Practices in Ireland
Chris Barry and Debbie Charleton

SHORT PAPERS

AN OPTIMIZATION METHOD FOR REDEMPTION AND DUE DATE MATCHING IN ASSIGNMENT OF ELECTRONIC RECEIVABLES BY USING INTEGER LINEAR PROGRAMMING
Toshiyuki Moritsu and Norihisa Komoda

E-LEARNING FOR NEW GRADUATE EMPLOYEES - Another Function of e-Learning for New Graduate Employees of Japanese and Korean Companies
Jiro Usugami

OPERATIONAL AND BEHAVIOURAL DIMENSIONS OF E-SUPPLY CHAINS AMONG MALAYSIAN'S SMES
Kay Hooi Kee, Mohamed Zairi and Khalid Hafeez

A COMPARISON OF WEB SITE ADOPTION IN SMALL AND LARGE PORTUGUESE FIRMS
Tiago Oliveira and Maria F. O. Martins

PROTOCOL OF AUTHENTICITY TO PROVIDE LEGAL SECURITY IN E-CONTRACTS - A Prototype
João Fábio de Oliveira, Cinthia O. de A. Freitas and Altair Santin

USING TECHNOLOGY ACCEPTANCE MODEL TO EVALUATE USERS’ ATTITUDE AND INTENTION OF USES
Dauw-Song Zhu and Chih-Te Lin

SEMANTIC INTEROPERABILITY FOR E-BUSINESS IN THE ISP SERVICE DOMAIN
Jane Hall and Stefanos Koukoulas

ON-DEMAND MOBILE CRM APPLICATIONS FOR SOCIAL MARKETING - Business and Technology Perspective
Viktor Kaufman, Yuri Natchetoi and Vasily Ponomarev

E-COMPLEMENTARITY - The Link to e-Business Value
Pedro Soto-Acosta and Ángel L. Merino-Cerdán

A RESEARCH MODEL OF CUSTOMER RELATIONSHIP MANAGEMENT SYSTEMS FOR MOBILE DEVICES - Description of a Research Model about Customer Relationship Management Projects
Rebecca Bulander

POSTERS

A STUDY OF INNOVATION DIFFUSION OF ELECTRONIC PATIENT RECORDS FOR SUPPORTING MEDICAL PRACTICE
Vincent Cho and Geoffrey Lieu

MANAGEMENT INFORMATION SYSTEMS IN ROMANIAN UNIVERSITIES
Ana-Ramona Bologa, Razvan Bologa, Gheorghe Sabau and Mihaela Muntean

A FEATURE EXTRACTING METHOD FOR TAMPER DETECTION IN PRINTED DOCUMENTS
Yoshiyasu Takahashi, Takaaki Yamada and Seiichi Susaki
THE AFFECTION OF DEMOGRAPHIC CHARACTERISTICS ON MEDIA CHOICE - A Case about Chinese Miniature Automobile Consumers
Dao-ping Chen and Wei Liu

WEB BASED COLLABORATIVE DOCUMENT CREATION AND REVIEW SYSTEM
Marius Ioan Podean, Raluca Arba and Loredana Muresan

ANTECEDENCES AND CONSEQUENCES OF E-SERVICE QUALITY ACROSS INDUSTRY SECTORS
Daw-Song Zhu, Chih-Te Lin and Yu-Ling Su

RFID PASSWORD MANAGEMENT METHODS FOR FALSIFICATION PREVENTION IN BOOKSTORE MANAGEMENT USING SECURE RFID TAGS
Yuichi Kobayashi, Yoji Taniguchi, Toshiyuki Kawana and Masanori Akiyoshi

WORK LISTS FOR THE TRANSPORT OF PATIENTS - A Case for Mobile Applications in Health Care
Andreas Holzinger, Jürgen Trauner and Stefan Biffl

TECHNOLOGY VS BUSINESS NEEDS IN BUSINESS INTELLIGENCE PROJECTS
Ana-Ramona Bologa, Razvan Mărgăra and Adela Bara

WEBSITE INTERACTIVITY - e-Commerce Usability Perspectives in Indonesia
Vincent Didiek Wiet Aryanto

THE HONG KONG GOVERNMENT AUTOMATED PASSENGER CLEARANCE SYSTEM (E-CHANNEL) - A Study of Channel Management Strategies
Tak Ming Lam

THE PRIVATE AND PUBLIC PARTNERSHIP STRATEGY IN E-GOVERNMENT ESDLIFE IN HONG KONG
Tak Ming Lam

A WEB-BASED ARCHITECTURE FOR E-GOV APPLICATION DEVELOPMENT
Marcelo Tilli, André M. Panhan, Osman Lima and Leonardo S. Mendes

PROSPECTS OF GRID IN THE CURRENT VOTER REGISTRATION SCENARIO OF BANGLADESH
Sazia Mafiz

WEBSITE CREDIBILITY - A Proposal on an Evaluation Method for e-Commerce
Katsuya Watanabe, Masaya Ando and Noboru Sonehara

AUTHOR INDEX
INVITED
SPEAKERS
KEYNOTE
LECTURES
Keywords: e-Business, innovation, strategy, architecture, Internet, wireless, broadband, video.

Abstract: Innovation is invention or application of technologies or theories that radically alters business and the economy. For the last 200 years, innovation and the economy have been locked in 80-year cycles, which might imply that innovation is an economic driver, and vice versa. Based on this relationship, some forecast that innovation and the economy will decrease sharply due to several forces: a) rapidly decreasing economic growth, b) increasing demand for custom services, c) more entrepreneurial work environments, and d) urban and environmental degradation. Should such forecasts hold true, business may need to alter its offerings, operations and organization to survive. Such a scenario may also necessitate applied e-Business innovation: the combining of existing internet, wireless, broadband, and video technologies. One possible result: highly flexible front offices seamlessly integrated with highly efficient back offices. Such an e-Business could comprise: a) a customer-based and transaction-based organization, b) functions for adaptive offerings that anticipate consumer need, c) highly responsive, real-time, operations having no inventory, and d) value-based front-end, and automated back-end, decision making.

1 DEMOGRAPHIC CYCLES

The economy is complex but its change over time is not. A key to understanding its change is to focus on underlying forces. For example, demographic cycles can tell you if a market crash is just an extreme correction due to an overvaluation cycle or the beginning of a long-term economic decline. The economy appears to be cyclic: it over-expands when growth sets in and then cuts back to continue future growth. For example, since 1985, the U.S. economy has been growing due to rising earnings, spending, and productivity of the baby-boom generation. This demographic has driven economic growth (Figure 1) and corporate growth (Figure 3).

1.1 The Next Decade

Innovation is often associated with invention or application of technologies or theories that alter business and the economy. Innovation is a major economic driver, and the economy is a major innovation driver. For example, in the last 200 years, technology innovation and the economy were locked in upward 80-year cycles. Some forecast future U.S. economic downturn, due in part to the demographic cycle (i.e. a peak spending drop) and the end of the current 80-year technology innovation cycle (i.e. fewer new companies, products and jobs). Pundits see the cause for the downturn being primarily due to four forces: a) rapidly decreasing economic growth, b) increasing demand for personalized services, c) more entrepreneurial work settings, and d) urban and environment degradation.
1.2 The Birth Cycle

These forecasts are based on expected population growth peaks in many developed countries by 2010 and by 2070 worldwide. The rate of change is noteworthy: prosperity and urbanization appear to be causing rapid birth rate declines in industrialized countries (e.g. China, India, Europe, and Japan). The impact to e-Business innovation could be profound. Decreasing population growth over the next 30 years may cause less fundamental technology innovation (which occurs 25-35 years after birth) and less fundamental business innovation - decision making, organizational design, management theory (which occurs 45-60 years after birth) - in that same period.

1.3 The Spending Cycle

Spending is correlated to births. U.S. Department of Labor reports show two major components to U.S. spending. The first, the weaker of the two, is family formation, which occurs 25-35 years after birth (e.g. it drove the U.S. economy from 1955 to 1985). The second component, the stronger of the two, is peak spending, which occurs 45-60 years after birth and may drive the economy upward until 2010. At that time, population growth will slow, causing spending to slow likewise. This is causing some weakening in U.S. wireless sales, paid internet phone usage, search ad revenues, and online sales.

1.4 The Economic Cycle

Spending is coupled to innovation due to innovation creating new products and jobs. For example, fundamental (e.g. electricity, steel, and motors) and applied (e.g. assembly lines) innovation brought standard products to mass markets, along with factory jobs and urban living. Today, fundamental (e.g. wireless, internet, broadband, video) and applied (e.g. produce-to-order systems) innovation are bringing custom products to affluent markets, along with outsourced jobs and exurban living (e.g. small quality towns and distant urban rings). But, sustaining such remote work depends on today’s e-Business technologies and future innovation.

2 TECHNOLOGY CYCLES

Technology cycles and demographic cycles feed off each other: population grows and then innovation occurs, which enables more population growth. Five forces create a technology cycle: First, radical technology changes business fundamentals (e.g. railroads brought Sears goods to remote towns). Second, no one first knows how to profit from the new technology (e.g. 30-second TV ads). Third, shared infrastructure requiring large investment is needed (e.g. internet, wireless, broadband, video – see Figure 2). Fourth, the economy is healthy, and thus can make such investment (e.g. the U.S. economy from roughly 1980 to 2010). Fifth, low inflation favours investment in innovative firms. By the end of the cycle, a few companies survive (e.g. Dell, eBay, Amazon, AOL, Yahoo!, Cisco, Google).

2.1 The 80-Year Cycle

Major technology cycles last roughly 80 years and have four stages: Startup: Fundamental innovation causes new companies to emerge. Growth: Those firms grow into the main-stream. Shakeout: Slow growth and overexpansion cause a consolidation. Maturity: The surviving companies compete for final market share. For example, the last major U.S. technology cycle, computing technology (Figure 2), began around 1950 with the advent of the early mainframes. Cheaper computers, fast networks, and massive storage emerged in 1995. Internet, wireless, broadband and video will reach maturity by 2030.

Figure 2: The Last Major 80-Year Technology Cycle.

Approximated market adoption s-curves:
1=mainframe, 2=minicomputer, 3=personal computers, 4=Internet, 5=wireless, 6=broadband.

The right y-axis is logarithmic with two unit sets:
S=storage capacity in gigabytes (GB), and
N=network capacity in bites-per-second (b/s).
2.2 The Internet S-Curve

It took about 20 years for the Internet to have widely accepted standards\(^\text{13}\). Similarly, cars reached 10% of U.S. homes in 1914, and then jumped to 90% market penetration by 1928, with a shakeout occurring in 1921. Likewise, the Internet reached 10% of U.S. homes in 1996, grew fast, then hit 50% penetration in 2001. Near that point, a major industry shakeout was expected and did occur – the “Dot Com Crash” happened about 80 years after the 1921 automobile industry crash. Internet usage has now reached its maturity stage, and firms such as eBay, Amazon, AOL and Google made it through the shakeout.

2.3 The Wireless S-Curve

Technology adoption follows an S-curve pattern: a new technology goes “main stream” and then grows fast, much faster than the economy, until it reaches 90% market penetration. Wireless technology (e.g. mobile phones, PDAs) penetrated U.S. markets on pace with Internet adoption, and now electronic commerce services are being offered to end-users\(^\text{14}\). It hit 10% market penetration in 1994, 50% in 2001, and hit 90% this year. Most importantly, the combination of internet, wireless, broadband and video technologies is being looked upon by many as one likely next wave of e-Business innovation\(^\text{15}\).

2.4 The Broadband S-Curve

So, a major change in e-Business is possible, and it may occur as broadband and video technologies reach most individuals. Broadband connections (e.g. DSL, cable modems) hit 10% market penetration in 2001, over 30% in 2004, and the shakeout is formally over. Digital cameras and wi-fi networks are emerging at similar rates. By 2030, the combined technologies of internet, wireless and broadband will reach 90% market penetration. This maturity, coupled with multi-modal e-Business transactions\(^\text{16}\), sets the stage for applying technology combinations to support affluent and niche markets.

3 CONTEXT AWARE BUSINESS

Economic and social success drives the rise-and-fall business cycle. Each rise calls business to reorganize its structures for higher population, wealth, and standard of living. This occurred before 1914, when fundamental technology innovation (electricity, steel motors) enabled the assembly line, followed by fundamental business innovation (Sloan’s product divisions and functional units) that enabled office work and suburbs. The coming period of decreased spending may again require business to reorganize – this time to become hyper-aware of customers, and the socio-economic factors that affect their buying\(^\text{16}\).

3.1 Arrival of Mass Affluence

Mass affluence is the current U.S. economy\(^\text{3}\), where “affluence” is defined as a household having income over $100,000 and net worth over $500,000, apart from the home. In 2001, there were 20 million such households, with 30 million expected by 2009. The latter may account for 50% of total spending. This affluent market is not wildly wealthy, but is beyond the middle-class standard of living that emerged in the last economic cycle. As it ages and becomes austere, this group could dominate U.S. markets, and could reshape business for several decades (e.g. demanding premium products at value prices\(^\text{104}\).

3.2 Premium Market Growth

From 1970 to 1990, discount firms (e.g. Wal-Mart) made goods very affordable (Figure 3). This freed up discretionary income for premium goods (e.g. Callaway). So, standard firms (e.g. Sears) are getting squeezed out. But the expected austerity wave may create a new value market segment, where aging consumers obtain premium products at discount prices by using electronic bargaining agents\(^\text{92}\). In such a market, an e-Business could exploit XML to define and create product fragments that can later be combined into highly customizable solutions\(^\text{17}\).
3.3 Old Culture Scrutinized

Firms are now recognizing mass affluence. Affluent households: a) value quality over quantity, b) value service over price, c) make their own decisions, and d) make a difference versus just doing a job. This means new consumption, and a new business model (e.g. individual pricing for mobile e-Commerce services). If the old culture (devalue customer time, optimize worker time, suppress worker talent and motivation, and maximize shareholder return while service levels drop) is scrutinized, that may set the stage for a corporate power shift.

3.4 Corporate Power Shifts

While at General Motors, Alfred Sloan invented a new corporate model that gave trade-up brands to the middle class. But this model is now meeting diminishing returns in the face of premium product growth. Rising self-esteem and self-actualization of workers are causing them to exit to start their own business or to work for higher-growth firms. A new business model – producing personalized products – is emerging that can give a competitive edge to those firms that understand and implement that design early in the next major 80-year economic cycle.

3.5 New Management Model

Personalized products require many fragments and combination options. The old business model has too many top-down policies trying to coordinate too many processes to allow for personalized products or service. The solution is the produce-to-order model. It coordinates real-time production through the automation of logistical and scheduling tasks. It permits direct ordering and delivery of customized goods with “little to no” inventories and less bureaucracy – Dell being one example. Companies that embrace this new model can use e-Business concepts and constructs to reorganize into a network of smaller businesses, subcontractors and vendors.

4 CUSTOMER-BASED ORGANIZATION

The produce-to-order model, with its bottom-up management powered by internet, wireless and broadband, leverages software, data and networks to maximize business response and productivity. Wintel technology that created the client-server platform for distributed office work, has given way to the instantaneous, network-centric, World Wide Web. But e-Business 2.0 cannot be fully realized without simplifying the front-end, and that occurs with decentralized decision making. Thus, complex front-end processes, and back-end bureaucracy, must be simultaneously simplified (e.g. Figure 4).

4.1 End-to-End Transaction

When e-Business is designed top-down, it leaves a firm incapable of defining and executing the optimal response to the customer event that initiated a transaction. “Produce-to-order” lets the e-Business focus on the entire, end-to-end transaction from the customer’s perspective. The result: the spanning of all parties, activities, events, responses, messages and data across the entire supply chain. For example, multiple firms participating in a federated execution of a single transaction require: a) coordinated and secure message flows, b) information exchange agreements, and c) coordinated message tracking.

4.2 Customer-Based Business Design

Satisfying an affluent market means adapting the e-Business for each customer and small market. This means personalizing each front-end sub-transaction according to the customer’s language, culture, etc. This is context-sensitive design. The context is: customer, events, and expected responses. The design is what some now call “customer-based inter-organizational systems” and some call it activity building authenticity into operations. As a design aid, models can be developed to explain how customers could use an e-Business, and such models have been shown to help people and systems adjust operations to accommodate customer expectations.

---

Figure 4: A Customer-Based Organization.
4.3 Business Replication by Market

All customers are not alike; and now, using current internet technology, like-minded customers can band together via electronic brokers, which bargain on their behalf to acquire custom products at value prices. The result is a plethora of dynamic small market segments! So, the front-end sub-transactions should be built on an adaptable e-Business platform that enables rapid replication and alteration to fit the language, culture and nuances of each customer or market segment. Otherwise, an e-Business may not keep up with customers changing their events and expectations around the responses to those events.

4.4 Full Service Response

To achieve high adaptability, the response to an end-to-end transaction should come from a distinct functional entity that is standardized to yield profit. A best-in-class design has: a) highly modular, plug-and-play, responses, b) a public interface of defined events and responses, and c) dynamic, context-sensitive, class loading of non-standard responses. This “full service response” comprises modular, coupled, and optimized sub-transactions. It is object-oriented with functionally-oriented, standard components, each of which is designed to achieve the same level of: reputation, user trust, information quality, functional availability and readiness, speed of response, and domain-specific characteristics.

4.5 Browsers and Butlers

Personalization increases product complexity and information distribution. Thus, human assistants, narrowly focused “browsers,” are needed to help customers choose, personalize, and use products. These e-Professionals oversee: trust, security, privacy, version and access control, configuration management and delivery dispersion. Similarly, back-end assistants are needed, since similar buyers can have conflicting priorities. These “butlers” have a wide focus: they know the whole end-to-end transaction in depth. This is crucial; service failures are directly related to a lack of in-depth knowledge, resulting in product/service personalization errors.

5 ADAPTIVE SOLUTIONS

The Internet is an adaptive medium. Its mechanisms can change a product, a service, or a brand faster than other media. It can distribute changes almost instantly across a company, a market, or a supply chain. It is an ideal platform upon which a solution can be structured, configured, delivered and serviced to meet affluent buyers who demand personalization. This section and Figure 5 give an example of how an adaptive service solution can be architected.

5.1 Value Chain

The increasing solution complexity that is now driving buyer-supplier relations in consumer-based markets is addressed in three stages. First, all companies that touch an end-to-end transaction are organized into a value chain. Second, that transactional organization is made to operate on an internet-based platform that can: a) be rapidly branded, and b) selectively opened or closed to any customer or market segment. Third, ebXML (i.e., today’s defacto standard for message exchange, trading protocol, common terminology, and registered process) is used by each company in the value chain to implement its standard response.

5.2 Internet Branding

Each value chain member has a brand. Highly personalized solutions require the e-Business to: a) preserve brand equity for each member, b) enable brand change, emphasis and transparency, and c) enable member differentiation, identification and interactivity. For example, individuals are part of the value chain for a contract labour solution. Today, people brand themselves on social networks. The staffing industry calls this a “video resume.” This language falls short: a resume documents experience and skills, whereas social networks can demonstrate competency, demeanor, presence, and articulation.

Figure 5: A Service Solution with Real-Time Branding.
5.3 Brand Transparency

During design, each internet brand goes through a chartering process: create, structure, communicate, direct, manage and maintain. “Transparency” is the decision to: a) hide who is responding to achieve transactional continuity, or b) show who is responding to build trust. The design is implemented using internet frames to “nest” brands within brands. For example, the checkout sub-transaction of an online purchase may make visible the checkout vendor’s brand or logo to build trust with the buyer.

5.4 Configurable Workflows

When an e-Business is a service, it comprises many sub-transactions, each having an operational life of its own. One can describe each sub-transaction by a workflow. For adaptability, all workflows should be highly modular and granular, with well-defined configuration rules. This enables reconfiguration of workflows in real-time to meet customer preferences and business standards. For example, today, artificial intelligence technology is used to deduce customer need, buying intent, and tendencies. This knowledge is then used to reconfigure the workflows for each specific customer or small market segment.

5.5 Data Integrity and Privacy

Data integrity across workflows permits flawless end-to-end transactions. Failure or delay occurs if sub-transactions use different data definitions or if they shirk their responsibility. Such cases can be mitigated using secure message-oriented middleware having a shared data object pool, and passing extra messages to verify data and function alignment. In line fashion, data privacy should be built-in: By isolating the back-end from the front-end, response data is protected by limiting export to the functions needed for a response. Data privacy must include a policy, automated audits, and formal consent.

5.6 Anticipation of Customer Need

Reengineering is not anticipatory. Instead, back-end butlers study analytical reports of customer activity to understand collective needs, and use that knowledge to migrate standard workflows to better serve all customers. At the same time, front-end browsers are building trust and reputation with customers and the value chain. Once in a trusted position, they participate in strategic-level dialog for new uses of the value chain. They create rational trust by: a) engaging with the customer, b) listening and framing needs, c) envisioning new solutions.

6 RESPONSIVE OPERATIONS

Moving to a transactional organization that produces adaptive solutions is a migratory activity. Success starts with mapping current operations: organization, codes, process, rules, and so on. Using that picture, operational elements are identified and modularized into autonomous business functions that correctly contribute to each and every response to each and every business event, and any sub-transaction that initial event spawns. This section and Figure 6 give an example of the result of this design activity: a full service response having specific operational goals: order and service management, sourcing strategy, transactional alignment, and real-time reporting.

6.1 Order Management

Orders actually start when customer and provider share organizational knowledge and XML forms for orders and fulfillment. The former comprises codes (e.g. departments), structure (e.g. hierarchy), and rules (e.g. designees). Once this knowledge is shared, optimal ordering can be achieved using electronic flow down of orders and fulfillment rules. This kind of digital binding of companies using a single management scheme transforms operations and reduces variations and handoffs. Note here that “optimal” does not mean “minimal.” In a complex value chain, some suppliers become inefficient when they are electronically integrated.

[Diagram of Adaptive Architecture for Operational Response]
6.2 Sourcing Strategy

Sourcing is complex, especially when opportunity timing is a deciding factor. Some sourcing can be automated by: a) locating opportunities via strategy-based profiles, b) adjusting selection criteria using prior query experience, c) maintaining opportunity classifications, and d) using analytics to refine that classification. Best-in-class profiles have three dimensions: differentiation, cost, personalization. Many tools now exist to "mine" internet sources, and their effectiveness hinges on building trust. But, companies still rely on manual sourcing when: a) supplier trust is unknown, b) duties, taxes or quotas are required, c) logistics or transportation is involved, d) transaction risks exist. All these factors go into an internet sourcing strategy.

6.3 Service Management

Uncontrolled product personalization (during sales) leads to inaccurate demand forecasts, high inventory investment, and poor customer service. So, just like with product replication by market, the variations of service delivery must be carefully controlled. Service management after product delivery is thus crucial. For example, a contract labour management service is often implemented along two dimensions: the actual work of the hired person, and the human resources management of that person. When service management succeeds, the tactical objective of the value chain solution is met: well-managed orders; fulfillment within customer-desired service levels; satisfactory quality of the delivered commodity; good follow-on service. Tactical fulfillment paves the way for the provider to become a trusted advisor.

6.4 Transactional Alignment

Nothing can ruin response faster than transactional misalignment. Response requires sub-transactions to work correctly, and correctly together. Misalignment causes process, product, or data errors – the results of which require costly adjustments after transaction closure. For example, incorrect state tax on labour, an error found in contract labour solutions, creates charges which require subsequent adjustment. These can be eliminated if all sub-transactions share the same tax rules. So, the design or deployment of an e-Business should be reviewed end-to-end to ensure function, data, and rule alignment among all of its sub-transactions.

6.5 Real-Time Reporting

A best-in-class e-Business has three types of reports: Progress Reports tell how the value chain is doing against the customer's outsourcing objectives. Best-in-class versions have two dimensions: product or service quality, and delivery process. Predictive Reports visualize order patterns and deduce their causes: need, business cycle, economic context, etc. Best-in-class versions anticipate customer need and predict if that need can be met. Performance Reports trend defect, compliance, productivity, etc. Best-in-class versions use a variety of performance metrics, capture highly granular data, and display summaries and trends via desktop dashboards.

7 OPTIMAL DECISION MAKING

A transactional organization becomes virtual when multiple companies participate in the e-Business. Optimal decision making happens if the e-Business permits autonomy, cooperation and control among business functions. Today’s top companies use this approach to adjust their operations to generate new electronic revenue streams and enable new service mix strategies. But since functions need rules to operate and coordinate, a transactional organization becomes durable only with a flexible framework of rules; specifically an XML-based framework for optimal decision making. This framework has two dimensions: standards-oriented (industry, company, customer), process-oriented (language, practices, service). This creates a design space (e.g. Figure 7a, b, and c) for XML-based code sets that together, can support value-based front-end and automated back-end decision-making for an e-Business.
7.1 Language-based Decision Rules

The semantic automation of an e-Business starts with industry-specific terminology that is captured in an ontology. This framework comprises five distinct sets of terms: a) process, b) information, c) application, d) data, and e) infrastructure. Optimal operations can occur when terms are standardized. But language must be a strategic consideration; a shared company language is insufficient unless it accounts for nationalistic interpretations of words and their implied meanings. Within the context of intensified competition in a market that demands high personalization, the ability to market, sell, and support using the customer's language becomes imperative. So, an e-Business architecture should contain: a) a message-passing mechanism for buyer-supplier interactions, and b) a declarative language for expressing customer requests.

<table>
<thead>
<tr>
<th>Interface</th>
<th>Scope</th>
<th>Contents</th>
<th>How Implemented</th>
</tr>
</thead>
<tbody>
<tr>
<td>Input</td>
<td>What</td>
<td>Data Definitions</td>
<td>XML, XML Schema</td>
</tr>
<tr>
<td>Output</td>
<td>What</td>
<td>Data Definitions</td>
<td>XML, XML Schema</td>
</tr>
<tr>
<td>Control</td>
<td>When</td>
<td>Events, Triggers, Coordination</td>
<td>XML, XML, ebXML</td>
</tr>
<tr>
<td>Mechanism</td>
<td>How</td>
<td>Algorithms, Standards, Regulations</td>
<td>ebXML</td>
</tr>
</tbody>
</table>

Figure 7b: Process-Oriented Interface Definitions.

7.2 Practice-based Decision Rules

Once language is settled upon, standard practices can emerge. At the highest level, these are industry best practices. When encoded in XML, including measurable objectives and dependency relationships among practices, they can easily be embedded into manuals or systems. The second level of practices, often called standard operating procedures (SOP), simultaneously enables business replication by market and full service response. When encoded in XML, this company knowledge can be immediately deployed exactly where and when needed. The third level of practices, used during replication by market, adjusts the SOP to meet customer needs (e.g. market niche, natural business cycle, unique value proposition, special contracts). The best way to adjust is to overlay customer operations atop the SOP. This way, company functions are not changed, but instead are replaced, by customer functions.

<table>
<thead>
<tr>
<th>Interface</th>
<th>Focus</th>
<th>Industry Standard</th>
<th>Company Standard</th>
<th>Customer Required</th>
</tr>
</thead>
<tbody>
<tr>
<td>Input</td>
<td>Language</td>
<td>Ontology</td>
<td>Meta Data</td>
<td>Meta Data</td>
</tr>
<tr>
<td>Output</td>
<td>Language</td>
<td>Ontology</td>
<td>Meta Data</td>
<td>Meta Data</td>
</tr>
<tr>
<td>Control</td>
<td>Service</td>
<td>Benchmarks</td>
<td>Company SLAs</td>
<td>Customer SLAs</td>
</tr>
<tr>
<td>Mechanism</td>
<td>Practice</td>
<td>Best Practices</td>
<td>Standard Operating Procedure</td>
<td>Customer Functions</td>
</tr>
</tbody>
</table>

Figure 7c: Standards-Oriented Rule Set Overlays.

7.3 Service-based Decision Rules

With language and practices in place, service levels are then defined. The first set comprises an industry benchmark for each sub-transaction. When added together, a true measure of "response" is obtained for each end-to-end transaction. These metrics feed the performance reporting system. Inside industry benchmarks are metrics for SOP. Besides speed and cost, they can measure usability, trust, loyalty, innovation, flexibility and financial impact, and they can occur when the buyer, or the supplier, or both, are mobile. These metrics feed the predictive reporting system. If the customer requires service levels that exceed industry benchmarks and SOP, the customer-facing business functions receive new metrics. Besides speed and cost, service levels can define product quality and service quality. These metrics feed the progress reporting system.

8 SUMMARY AND CONCLUSIONS

e-Business is here to stay, but it may likely change over the next 15 years as population growth declines impact innovation and how internet, wireless, broadband and video technologies are combined and used to conduct e-Business. One possible future scenario is the increase in applied innovation, built on new XML-based architectures, for providing highly flexible front-end and highly standard back-end e-Business platforms. To summarize:

8.1 More Applied Innovation

Innovation is invention or application of theories or technologies that radically alters business and the economy. Since innovation is a major economic driver, and vice versa, any economic downturn could impact e-Business innovation. Slowing population...
growth and exiting aged workforces can deplete both the number of inventors and those capable of using investments for fundamental business innovation. In parallel, the simultaneous maturation of the Internet, wireless, broadband and video may provide new opportunities for combining these technologies into solutions for the mass affluence economy. More applied, and less fundamental, innovation is likely.

8.2 Architecture Innovation

Those consumers capable of demanding premium, personalized products dominate the mass affluence economy. These consumers are also tech-savvy, and can exploit the Internet by creating intelligent agents that build dynamic small markets of like-minded consumers to negotiate with suppliers for premium, personalized products at value prices! Since this market could grow to over 50% of the overall economy, e-Business must become hyper-aware of their needs and buying patterns, and must be “architected” into a highly flexible front-end (to flex to each individual or small market) and a highly standardized back-end (to enable cost-effective operations). Architecture innovation could be a key to e-Business survival in the coming decades.

8.3 XML Platform of Contracts

Internet technology now exists to “architect” flexibility with standardization. A crucial technology is XML. It is an internet-based software language for defining business language, practices and service levels (thus enabling standardization), and allowing them to be easily changed and distributed across all business functions (thus enabling flexibility). Couple this mechanism with an architecture of “customer within company within industry,” and you have a platform for standardizing business contracts and then overlaying the more esoteric rules required for meeting the varied and changing demands of the individuals and small markets in the mass affluence economy. Future market forces could likely make an XML-based contracts platform commonplace.

8.4 Front-Line Decision-Making

Such architecture requires an investment – shed top-down control, reduce bureaucracy and decentralize decision-making. The XML platform of contracts lets a company define its operational rules while also enabling giving up control for how to apply those rules for each customer or small market segment. That is the job of front-line “browsers,” who know their customer best. They can apply standard rules appropriately, and can overlay customer-specific rules atop the corporate standard. At the same time, back-end “butlers” use their “all customer” perspective to improve the standard rule set. In this way, best practices are centrally defined and immediately distributed to the front lines. Context-sensitive e-Business (i.e. high-flex front-ends with standard back-ends) may become the norm.

ACKNOWLEDGEMENTS

Thanks go to Mr. Harry S. Dent (www.hsdent.com) for his research on the interrelationship among population growth, the economy, and technology innovation, and to Dr. Clement McGowan for his insightful review of earlier drafts of this paper.

REFERENCES


BRIEF BIOGRAPHY

David A. Marca is on the Adjunct Faculty at the University of Phoenix. His six books and 24 papers cover e-Business, e-Commerce, business process reengineering, and software engineering. He holds a
patent in workflow technology. His last book, entitled “Open Process Frameworks: Patterns for the Adaptive e-Enterprise,” was published by the IEEE in 2006. David is also President of OpenProcess, Inc. – an e-Business consulting firm since 1997 – that helps firms implement workforce management and e-Business solutions. He has consulted in Italy, Norway, Mexico and the United States. David is a member of the Institute of Electrical and Electronic Engineers (IEEE), the Association for Computing Machinery (ACM), and the Project Management Institute (PMI).
IMPROVING RELIABILITY IN COMMERCIAL IP NETWORKS

Yaakov Kogan
AT&T Labs, 200 Laurel Avenue, Middletown, NJ, U.S.A.
yakovkogan@att.com

Keywords: IP network, IP router, availability, reliability, redundancy.

Abstract: As the Internet becomes an increasingly critical communication infrastructure for business, education and civil society in general, the need to understand and systematically analyze its reliability becomes progressively more important. An Internet Service Provider (ISP) faces a challenge of providing service that meets customer expectations in terms of price and reliability while continuously reducing its cost. In fact, changes in Internet technology (particularly software) are significantly more frequent and less rigorously tested than used to be in circuit-switching telephone networks. An ISP can wait until the technology will mature but then it faces the risk of losing customers and revenues in a situation where many customers care more about low prices than claims of better quality and reliability. A large ISP has to meet high reliability requirements for critical applications like financial transactions, Voice over IP and IPTV. This results in variety of redundancy solutions at the edge and resilient core which is shared by traffic from all applications. In this paper we review redundancy solutions for eliminating the customer impact in typical failure modes in IP networks and present a constructive metric for evaluating the reliability of commercial IP networks.

1 INTRODUCTION

The ARECI Study (Bell Labs, 2007) conducted for the European Commission states “European security, economic stability, and the public safety and welfare of its citizens increasingly depend on the availability and robustness of its electronic communications infrastructures”. The main focus of the Study is the crucial subject of the end-to-end availability of communication networks. It emphasizes that “European citizens are used to the high reliability of legacy telephone service and come to expect new services (e.g., VoIP, Internet, IPTV) to have similar level of reliability. Legacy telephone service uses circuit switching, allocating an exclusive path with fixed capacity across the network for the duration of the call regardless of the actual usage of the circuit capacity. The reliability objective of ‘no more than 2 hours downtime in 40 years’ has been applied to voice communication since 1964 (Malec, 1998). It has been achieved using expensive redundancy solutions for both switches and transmission facilities. Commercial IP networks have three main advantages in comparison with legacy telephone networks. First, packet switching as the transport mechanism provides far more economically and technically efficient multiplexing of network resources. Second, protocols like MPLS TE with FRR are more economically efficient in restoration from failures of transmission facilities than traditional 1+1 redundancy. Third, commercial IP networks can provide different level of redundancy to multiple services and price the service dependent on its reliability. That allows Internet service providers to satisfy many customers who care more about low price than reliability (claffy, Meinrath and Bradner, 2007) and recover high cost of redundancy with higher price for mission critical applications.

This paper describes redundancy solutions for eliminating the customer impact in typical failure modes in IP networks and provides the framework for quantification of reliability for commercial IP networks.

2 REDUNDANCY SOLUTIONS

Redundant elements in IP routers and networks are necessary to provide service restoration without human intervention. However special mechanisms and protocols are required to reduce restoration time...
to sub-second level (required for video Tele-
Presence) and further below 100 milliseconds
(required for broadcast video and some financial
applications). The total restoration time in IP
network can be broken up into the following three
phases:

- Detection: time required to detect and react to a
  failure;

- Notification: time required to communicate the
  failure in the previous phase to all relevant
devices in the network;

- Restoration: time required to move traffic to an
  alternate network device or path.

In this section we list key redundancy solutions
and protocols which eliminate service interruptions
that can be caused by router and facility outages.

- Non Stop Routing (NSR) provides hitless
  failover to backup Route Processor when
  primary RP fails. It eliminates simultaneous
  failure of all line cards in case of RP failure.
  NSR is essential for achieving the required
  MTBO level for the Backbone network. NSR is
  the first building block for In Service Software
  Upgrade (ISSU) that must eliminate or
  significantly reduce planned outages due to
  software upgrade.

- Fast Reroute (FRR) is a feature in an MPLS
  network that enables the service provider to
  provide a redundant data path should a link
  (facility) or router line card fail. FRR is
  intended to offer practically hitless restoration
  with failover times on the order of tens of
  milliseconds. To achieve such a fast restoration
time, a backup Label Switched Path (LSP) is
  established at each node. The traffic is switched
to the backup LSP once a failure has been
  detected in the primary LSP. Typical budget for
  achieving 50 milliseconds for the total
  restoration time is 10 milliseconds for Detection
  and Notification and 40 milliseconds for
  Restoration (traffic redirection). Today only
  facility failures can be detected and notified
  within 10 milliseconds. Reduction of detection
  and notification time for other failures is work
  in progress.

- Bidirectional Forwarding Detection (BFD)
  provides fast failure detection time (on the order
  of tens of milliseconds) between customer and
  access routers. BFD can be used to trigger
  reroute to the backup access router if the
  primary access router fails.

3 THE FRAMEWORK FOR
RELIABILITY ANALYSIS

For reliability analysis, IP network consists of
routers interconnected by transmission facilities.
Routers are responsible for packet delivery to the
destination while facilities provide the necessary
bandwidth to avoid packet loss. An ISP typically
sells to customers access ports. These ports are
housed on Line Cards that are on the “drop side” of
an Access Router, where facilities from a Customer
Router terminate.

There are five element types in a typical IP
Access Network topology whose failure can cause
downtime for customer facing Line Cards:

- Line Card on the customer facing side. Any
  failure in the electronic or optical components
  of the Line Card will result in Line Card
downtime.

- Access Routers that form an edge on an ISP
  Backbone network. A total router failure will
  result in downtime for all Line Cards

- Facilities and supporting elements such as
  cross-connects, which link Access Routers to
  Backbone Routers. To increase the reliability of
  the Access Network, an ISP usually provides
  redundancy by connecting each Access Router
to two Backbone Routers at the same access
node using two independent sets of uplinks.
This permits customer traffic to enter the
Backbone in the following failure scenarios:

  - A failed uplink
  
  - A failed card supporting an uplink
  
  - A failed Backbone Router at the access node

If all facilities linking an Access Router to a
Backbone Router fail, then all Line Cards at
the Access Router will experience downtime.
Backbone Routers linked to Access Routers. If both Backbone Routers at an access node fail (a rare event) then all Line Cards on the Access Routers at this node lose connection to the Backbone.

Facilities linking Backbone Routers at an access node, to Backbone Routers at other Backbone nodes. Such facility failures decrease the available bandwidth from Access Routers to the Backbone. Note that if all Backbone Router uplinks at an access node fail (a rare event), then all Line Cards on the Access Routers at this node lose connection to the Backbone.

Thus, the access Line Card acts as a common denominator for all of the above failure types. Any one of these failures results in downtime for the impacted Line Cards. In addition, failures in the Backbone without hitless restoration may interrupt traffic between many access Line Cards.

Let \( N \) be the total number of access (customer facing) Line Cards and traffic were impacted on \( K \) of these card during a fixed time interval \( T \) (e.g., one month measured in hours). Then Mean Time between Outages (MTBO) is defined as

\[
MTBO = \frac{NT}{K} \quad (1)
\]

Table 1 demonstrates application of MTBO for calculation of the probability of successful failover (Coverage) in the Backbone required to achieve the End-to-End MTBO of 40 years. Assume that MTBO for an Access Router without redundancy is 7 years. This case is listed in Table 1 as ‘Single Access’.

<table>
<thead>
<tr>
<th>Connection</th>
<th>Coverage</th>
<th>MTBO, years</th>
</tr>
</thead>
<tbody>
<tr>
<td>Single Access</td>
<td>96%</td>
<td>7</td>
</tr>
<tr>
<td>Dual Access</td>
<td>96%</td>
<td>172</td>
</tr>
<tr>
<td>Both Ends</td>
<td></td>
<td>86</td>
</tr>
<tr>
<td>Backbone</td>
<td>85%</td>
<td>75</td>
</tr>
<tr>
<td>End-to-End</td>
<td></td>
<td>40</td>
</tr>
</tbody>
</table>

The probability of successful failover to the backup device (e.g., router) when primary device fails is referred to as Coverage. Row ‘Dual Access’ in Table 1 shows that MTBO can be dramatically increased by providing dual homing, where customer router is connected to two access routers located in different Backbone nodes if we assume that the respective Coverage is 96%. The next row in Table 1 takes into account ‘another end’ of the network with similar two diverse access routers. To obtain the desired End-to-End MTBO of 40 years we need to allocate 75 years to the Backbone MTBO. Assume that mean time between failures of Backbone facilities or router cards is 2000 hours, and that, on the average, one such failure without hitless restoration affects traffic on 2% of customer facing Line Cards. Then the Backbone coverage is 85% which can be realistically achieved with NSR and FRR described in the previous section.

REFERENCES


BRIEF BIOGRAPHY

Yaakov Kogan received the Ph.D. and Doctor of Sciences degrees from the USSR Academy of Sciences in 1969 and 1987, respectively. He joined the Network Design and Performance Analysis Department at AT&T in 1993 where he is currently a Lead Member of Technical Staff. During 1989 – 1993 he was a professor at the Faculty of Industrial Engineering and Management of the Technion University. After receiving the Ph.D. he worked on performance analysis and measurements of computer and communication systems and developed nonparametric and asymptotic methods for solving stochastic models of large dimension. He published 5 books and more than 100 papers. Dr. Kogan is a member of IFIP Working Group 7.3 on Computer Systems Modeling since 1979 and IEEE Fellow since 2001. His recent activities include performance and reliability analysis of large IP and Frame Relay networks.
CRYPTOGRAPHIC ALGORITHMS

Successes, Failures and Challenges

Bart Preneel

Katholieke Universiteit Leuven, Dept. Electrical Engineering-ESAT/COSIC
Kasteelpark Arenberg 10, bus 2446, B-3001 Leuven, Belgium
IBBT, Van Crommenlaan, B-9000 Gent, Belgium
bart.preneel@esat.kuleuven.be

Keywords: Cryptology, information security.

Abstract: The exponential progress of hardware during the past decades and the explosion of wired and wireless networks has resulted in a large scale deployment of cryptography in financial applications and electronic commerce, in mobile phones and in electronic identity cards. In parallel with his deployment the insight has grown that building secure information systems is a very hard problem: cryptography is an essential building block that may have helped us to win some battles, but one still has the impression that we are losing the information security “war.” This article clarifies the role of cryptology in information security. Subsequently it discusses the challenges that are faced by this discipline. There is a strong need for algorithms that offer better trade-offs between performance, cost and security. In addition we face challenges related to secure implementations in software and hardware and in the area of algorithm agility. We will conclude with a brief discussion of opportunities offered by the progress made in the area of cryptographic protocols.

1 INFORMATION SECURITY AND CRYPTOLOGY

Cryptology is the science that studies mathematical techniques to protect digital information (Menezes et al., 1997). It allows to identify entities in a reliable way, to establish secure communications over open networks, and to protect stored information. At a higher abstraction level, cryptology enables the collaboration of mutually distrusting parties towards achieving a common goal. Cryptology is essential for securing communications; one can think of mobile networks such as GSM and 3GSM, Virtual Private Networks, Wireless LANs and Personal Area Networks (e.g., Zigbee and Bluetooth). Cryptology also plays a key role in securing financial transactions; the financial world uses cryptology in its back-office, but also for identification in terminals (e.g. the EMV standard (EMV, 2004)), and for Internet banking. The TLS protocol in browsers is widely used for securing e-commerce transactions. There is also a growing use of cryptology for hard disk encryption; the Trusted Computing Group has introduced the TPM chip (Trusted Platform Module) that offers services such as secure key storage and remote attestation in tens of millions of computers. Most governments are introducing electronic identity cards or electronic health cards to offer secure on-line services to their citizens. One can anticipate that in a future world with trillions of electronic devices (known as wireless sensor networks, the Internet of things, ubiquitous or pervasive computing,...) we will see a further increase of cryptographic deployments.

During the last thirty years, cryptology has evolved from very limited use in restricted military and government environments towards a massive deployment. This development has been driven by the needs of the information society and has been enabled by the same development in microelectronics that has created this society in the first place. In parallel the science of cryptology has been progressed by building on complexity theory and information theory, in which security proofs (or more correctly security reductions) play an ever growing role. In this context the concepts are refined and we have substantially improved our understanding of the properties needed from the building blocks. Nevertheless, the development of many of these building blocks is still an engi-
engineering discipline where we learn and develop design principles by trial and error methods.

It is important to realize that in spite of the strong scientific track record, the cryptography problem is not solved. Cryptographic algorithms and protocols get frequently cryptanalyzed, because of progress in cryptanalysis and advances in computational power. Once the cryptography is broken, the security will probably collapse (e.g., decrypting sensitive information or copying controlled information becomes possible). This is very similar to a scenario in which a burglar would get access to a master key. In the best case one can use fraud detection and responsive technologies to control the losses, which would correspond in our building analogy to adding video surveillance and security guards. However, most security breaches are completely unrelated to cryptographic weaknesses (see for example Anderson (Anderson, 1994)). The most important reasons for these breaches are:

- incorrect or incomplete requirements for the system (this includes the case where a system is designed for one application and later on used for other applications with different security requirements);
- implementation errors (e.g., software bugs, cryptographic hardware or software with side channel weaknesses);
- configuration errors (this includes the case where algorithms and implementations known to be weak are not upgraded);
- human issues: the weakest link in any security system is humans; this has been illustrated by many war stories of hackers who use social engineering techniques to gain access.

The main reason why it is so difficult to make substantial progress is that our ICT infrastructure gets ever more complex: ICs have hundreds of millions of transistors, processors will have dozens of cores, operating systems consist of tens of millions of lines of code, applications are built from millions of lines of code and we have hundreds of millions of devices connected to the Internet. All these numbers will increase by another order of magnitude in the next decade. Securing complex systems is inherently difficult, and our approach to build simple and trustworthy subsystems with core cryptographic functionality is only partially effective. One of the future challenges is to learn how to isolate this functionality and how to integrate it with the rest of the system. We will also have to learn how to build resilient and adaptive systems with multiple lines of defense that degrade gracefully. A third area where we need to make substantial progress is understanding how humans interact with ICT systems and in particular how humans interact with security.

2 RESEARCH CHALLENGES FOR CRYPTOGRAPHIC ALGORITHMS

This section identifies two challenges for cryptographic algorithms: offering improved trade-offs between cost, performance and security and secure implementations.

2.1 Improved Trade-offs for Cryptographic Algorithms

Cryptographic algorithms offer a trade-off between cost (footprint in hardware or software, power and/or energy consumption), security and performance. Achieving any two of these three is straightforward provided that one can neglect the third one. It is trivial to make an insecure system that is fast and cheap. If cost is not an issue, it is rather easy to make a very fast and very secure solution. Similarly, if a system can be very slow, it can offer an acceptable security at a low cost. The real challenge is to improve these three parameters at the same time, in particular for specific extremes beyond the current state of the art, namely:

- **Extremely Low Cost Solutions** are essential to get cryptography everywhere (ambient intelligence, sensor networks and RFIDs). A specific target is encryption with less than 1500 gates or an entity authentication protocol that consumes less than 50 mJoules.

- **Extremely Fast Solutions** for applications such as bus encryption, and authenticated encryption for Petabyte storage devices and Terabit networks.

- **Need for High Security Solutions** for applications such as e-voting, e-health and national security we need cryptographic algorithms that provide guaranteed protection for 50 years or more. While this is conceivable for symmetric cryptography, this goal is currently a major challenge for public key cryptography, in view of progress in research to attack hard mathematical problems and the anticipated development of quantum computers.

This challenge will be illustrated by using as examples block ciphers, stream ciphers, cryptographic hash functions and public-key encryption schemes.
more detailed overview of recent progress in cryptanalysis can be found in (Preneel, 2007).

In the area of block ciphers, a major changed occurred when DES (a 1977 FIPS 46 standard designed by IBM (FIPS 46, 1977)) was replaced by AES (the 2001 FIPS 197 standard (FIPS 197, 2001)). The main shortcomings of DES are that its key size (56 bits) and block length (64 bits) are too small. M. Wiener has designed in 1993 a US$1 million hardware DES key search machine that can recover a 56-bit DES key in about three hours (Wiener, 1996). If such a machine would be built in 2008, it would be about 600 times faster, hence it would need about 10 seconds to find a key. The 64-bit block length is no longer adequate for most applications because there exist matching ciphertext attacks on the modes of operation of an n-bit block cipher which require about $2^n/2$ ciphertext blocks (Knudsen, 1994). In 1978, triple-DES was proposed as an alternative to DES; the two-key variant (with a 112-bit key) is believed to offer a security level of 70-80 bits, while the three-key variant (with a 168-bit key) offers a security level of 90-100 bits. Triple-DES still has a 64-bit block size. Moreover, DES is not very fast in software (45 cycles/byte on a Pentium-type processor), and triple-DES requires more than 100 cycles/byte. AES is the result of an open competition with as winner the Rijndael algorithm designed by Daemen and Rijmen (Daemen and Rijmen, 2001). It has a block length of 128 bits and key lengths of 128, 192 and 256 bits, which means that matching ciphertext attacks and exhaustive search are completely out of reach. Moreover, it is a very flexible cipher that offers a good software performance (15 cycles/byte), but also allows for a very broad range of size/speed trade-offs in hardware. Today AES is a trusted design that is being integrated in a growing number of applications; Intel has even announced that they plan to integrate a hardware implementation of one round of AES in their CPUs. For some low-end applications, a compact 64-bit block cipher with an 80-bit key would be useful; PRESENT (Bogdanov et al., 2007) is a recent design with these characteristics; it can be implemented in 1600 gates, compared to 3400 for AES.

The area of synchronous stream ciphers has a very long tradition that goes back to the 1960s (e.g., Golomb (Golomb, 1967)). Stream ciphers became less important after the introduction of the open block cipher standard DES, because DES was widely trusted and offers a flexible building block. However, today it is believed that stream ciphers still have the potential to outperform AES with a factor of two or more for very low footprint applications of for very high speed in software. Modern stream ciphers tend to move away from filter generators and combination generators based on Linear Feedback Shift Registers (LFSRs), since it turns out that these generators are vulnerable to fast correlation attacks (see e.g. (Meier and Staffelbach, 1989)) and algebraic attacks (see e.g. (Courtois and Meier)). Stream ciphers with clock control such as the shrinking generator (Coppersmith) and the self-shrinking generator (Meier and Staffelbach) seem to be more resistant to cryptanalysis; recently two new designs based on a Non-Linear Feedback Shift Register (NLFSR) have been introduced that have a very low footprint yet offer a good performance in hardware: Grain (Hell et al., 2005) and Trivium (De Cannière and Prennel). In software, the evolution seems to be towards designs that contain large tables that are updated at every iteration; RC4 is the pioneer of this area and recently HC (Wu, 2005) and PY (Biham and Seberry, 2007) have been added to this class. An alternative approach takes its inspiration from the block cipher designs (e.g., Scream (Halevi et al.), MUGI (Watanabe et al., 2004), and Salsa (Bernstein et al.). Between 2004 and 2008 the ECRYPT Network of Excellence has run the eSTREAM initiative with as goal to advance the state of the art in stream ciphers and to identify promising candidates for stream ciphers (ECRYPT). After an open call for proposals in 2004, 34 designs were submitted; during the evaluation phase more than half of these were broken. In April 2008 eight stream ciphers have been identified as promising candidates for future research and standardization (for more details see (Robshaw and Billet, 2008)).

In 2004, the hash function crisis started: Wang et al. showed that finding collisions for the widely used hash function MD5 (Rivest, 1992) requires only a few minutes on a PC rather than $2^{64}$ hash function calls; one year later they also showed that collisions for SHA-1 (FIPS 180-2, 2002) can be found in less than $2^{80}$ function evaluations needed for a brute force collision search (currently the best result announced by a team from T.U.Graz requires between $2^{69}$ and $2^{74}$ function evaluations). In parallel, several structural vulnerabilities were identified in our hash function designs, with perhaps as most surprising observation that the concatenation of the result of two iterated hash functions is only as secure as the strongest of the two (Joux, 2004). It should be pointed out that experts in the area of hash functions were perhaps less surprised by the results on MD5, since serious structural weaknesses had been defined in MD5 in 1993 by den Boer and Bosselaers (den Boer and Bosselaers) and in 1996 by Dobbertin (Dobbertin, 1996). A second observation is that the strength of MD5 and SHA-1 against brute force collision search is only $2^{64}$ and $2^{80}$; this means that brute force collision attacks
against MD5 with dedicated hardware were feasible in the 1990s; for SHA-1 a brute force attack will become feasible in the next years. One can conclude from this observation that MD5 should have been replaced more than 10 years ago. Many hash function designs have been broken; it seems that the designers are too optimistic: if MD5 would have had 8 rounds rather than 4 or if SHA-1 would have had 160 steps rather than 80, nobody would have spoken of a hash function crisis today. Due to their high speed in software and the freedom in usage (no license fees and less export issues), hash functions have become in the 1990s the Swiss army knifes of cryptography: they are used for a broad range of applications, including digital signatures, commitments, pseudo-random function, key derivation, identification protocols, construction of MAC algorithms, etc. The current crisis has generated an increased interest in the research community to understand the properties required for all these applications and to develop generic constructions. NIST has published in 2002 FIPS 180-2 (FIPS 180-2, 2002); this standard includes a.o. SHA-256 and SHA-512 and is commonly known as the SHA-2 family. While currently no viable attacks are known on this family, the similarity with the SHA-1 designs has created some concern. From 2008 to 2012, NIST organizes an open competition for the AHS (Advanced Hash Standard) or the SHA-3 family; the resulting standard will offer an alternative to the SHA-2 family.

In the area of public key encryption, all popular algorithms depend on a small set of problems in algebraic number theory (factoring (Rivest et al., 1978), discrete logarithm in $\mathbb{Z}_p$ (Diffie and Hellman) and discrete logarithm on an elliptic curve over a finite field (Avanzi et al., 2005)). If a breakthrough would be made in solving any of these problems, or if quantum computers could be built, we would have to abandon all these schemes. There exists a small number of alternatives based on coding theory and lattices. These alternative public-key schemes are known today under the name ‘post quantum cryptography’ since they could resist the existence of a large quantum computer. Typically they require more computation and/or memory than the schemes used today; in addition, more security analysis is needed before they can be widely adopted. In the last years several schemes based on multivariate polynomial equations have been cryptanalyzed. This shows that the design of novel public key encryption schemes is a challenging problem.

Today RSA is still the most widely used public key encryption algorithm. It is perhaps surprising to learn that it has taken cryptographers more than twenty years to understand what the security properties are for a public-key encryption algorithm and how one should use RSA in an application in such a way that breaking the security of the encryption can be reduced to breaking the RSA assumption. The most efficient construction is RSA-OAEP (Fujisaki et al.); note that it requires the random oracle assumption, which is somewhat controversial (for more details see Koblitz and Menezes (Koblitz and Menezes, 2007)).

### 2.2 Secure Implementations

The classified community realized already in the 1950s that in addition to secure algorithms, secure implementations are needed (see for example Wright’s book (Wright, 1987)). The open community discovered only in the 1990s the power of side channel attacks on software and hardware implementations of cryptographic algorithms. This includes passive attacks such as timing attacks (Kocher; Tsunoo et al., 2003), power analysis (Kocher et al.) and electromagnetic attacks (Gandolfi et al., 2001; Quisquater and Samide, 2001) and active attacks such as fault attacks (Boneh et al., 1997). By now it is clear that defending against these attacks is very hard, in particular since new and improved attacks are discovered on a regular basis. A broad range of countermeasures has been developed at all layers (circuit, logic gate, algorithms, protocols); very often they bring a substantial overhead and it happens regularly that countermeasures are shown to be insufficient. Researchers are also attempting to develop a theory to model side channel attacks in order to develop a solid foundation of this area. Overall, the development of implementations that are both secure and efficient is a very difficult challenge that will require a substantial research effort.

### 3 RESEARCH CHALLENGES FOR CRYPTOGRAPHIC PROTOCOLS

While this article focuses on cryptographic algorithms, there are also important challenges in cryptographic protocols. A first challenge has already been mentioned: the development of low cost entity authentication protocols for small devices. Substantial progress has been made in proving the security of simple building blocks such as authenticated encryption or two-party authenticated key agreement, but the development and analysis of more complex
cryptographic protocols for distributing trust is a major challenge. These protocols allow to reduce the requirement of trusting a centralized system and/or specific machines or hardware components by distributing this trust over a larger number of entities. This is very important for privacy sensitive applications such as voting, auctions and data mining (e.g., of medical data). The creation of complex cryptographic protocols has been based on an approach for provable security based on assumptions for cryptographic algorithms; due to the complexity of the proofs and methods, there is a need for advanced tools to develop and verify such proofs. A second goal is to take into account more realistic deployment models such as concurrent composition, asynchronous interactions, complex privacy models and protocols in which the players are not “malicious” or “honest” but rather “rational”; the latter approach results in a novel game theoretic approach.

A more practical problem is to design cryptographic protocols in such a way that algorithms or building blocks can be upgraded easily. Section 2 shows that the lifetimes of cryptographic algorithms are frequently shorter than anticipated. During the design of the SSL/TLS protocol, this was well understood, yet it turns out that MD5 and SHA-1 are part of the SSL negotiation process, and cannot be upgraded easily.

4 CONCLUSIONS

In this article we have not discussed the fundamental research questions in cryptology, that are also fundamental questions in computer science or mathematics, such as: can we prove that certain problems are hard, is it hard to factor the product of two large primes, or how difficult is it to solve large non-linear systems of equations, even if these have a regular structure (e.g., because they have been deduced from a block cipher). All of these problems are known to be very difficult, and researchers understand that so far we have only scratched the surface in this area, which forms the foundation of cryptography.

The goal of cryptography is to develop clear security definitions and develop an understanding of the relation between the security properties of the applications and those of the building blocks: algorithms (e.g., AES), primitives (e.g., the OCB mode based on AES, RSA-OAEP), protocols (e.g., an e-voting protocol) and applications. So far we seem to be making progress to prove relationships between protocols and primitives and between primitives and algorithms. However, we are still very far from analyzing complex applications and the security of cryptographic algorithms is to a large extent based on engineering and heuristics. For future systems, we should strive to create modular designs in which building blocks can be replaced as needed. There are also important challenges in the area of secure implementations; progress could be made here by developing models and methods to verify the security and correctness.

Finally, cryptography forms a building block of a larger security system. It is therefore important that we try to improve our understanding of the overall system requirements in order to ensure that the cryptographic solutions are as useful and usable as possible.

ACKNOWLEDGEMENTS

This article is based in part on the results of the ECRYPT project (ECRYPT, 2006). This author’s work was supported in part by the IAP Programme P6/26 BCRYPT of the Belgian State (Belgian Science Policy) and by the European Commission through the IST Programme under Contract IST-2002-507932 ECRYPT.

REFERENCES


ECRYPT eSTREAM initiative, http://www.ecrypt.eu.org/estream


**BRIEF BIOGRAPHY**

Bart Preneel received the Doctorate in Applied Sciences from the Katholieke Universiteit Leuven (Belgium) where he is currently a full professor. He was visiting professor at several universities in Europe. His main research interests are cryptography and information security. He has authored and co-authored more than 200 scientific publications. He is president of the IACR (International Association for Cryptologic Research) and a member of the Editorial Board of the Journal of Cryptology and of the IEEE Transactions on Forensics and Information Security. He has participated to 25 research projects sponsored by the European Commission, for five of these as project manager. He has been program chair of ten international conferences and he has been invited speaker at more than 30 conferences. In 2003, he has received the European Information Security Award in the area of academic research. He has been a member of the TCPA Advisory Board. He is president of L-SEC vzw. (Leuven Security Excellence Consortium), an association of 60 companies and research institutions in the area of e-security. He is cofounder and conductor of the jazz ensemble of the K.U.Leuven.
WATERMARKING, STEGANOGRAPHY AND CONTENT FORENSICS

Ingemar J. Cox
University College London, U.K.

Abstract: Electronic watermarking is about 60 years old. However it was not until the beginning of the early 1990's that watermarking received widespread interest, due to concerns about piracy of digital content. In the subsequent decade, very significant progress has been made both in our theoretical understanding of digital watermarking and in its applications. This progress is described here.

Steganography has a much longer history, dating back to at least the time of the ancient Greeks. While Shannon dismissed steganography as "{primarily a psychological problem!" , the last decade has seen the application of information theory to steganography. Terrorist events at the beginning of the 21st century motivated further attention and very interesting results have been described. At first sight, digital watermarking and steganography would appear to share the same goals. However, while both seek to hide information within other information or content, there are very significant differences in the constraints that must be satisfied. The similarities and differences between digital watermarking and steganography are highlighted here.

Steganography spawns steganalysis, the art and science of detecting the presence of a steganographic message hidden in innocuous content. Recent research views steganalysis as a binary classification problem; is a hidden message present or absent? Classification is based on testing for statistical anomalies in features derived from the content. Content forensics shares similarities with steganalysis. At the simplest level, content forensics is often asked whether, for example, an image is authentic or has been tampered with. This problem can also be viewed as a binary classification problem and similar techniques can be applied. We will review recent work in content forensics and steganalysis, and discuss the limitations of both.

BRIEF BIOGRAPHY

Ingemar J. Cox is currently Professor and BT Chair of Communications in the Departments of Computer Science, and Electrical and Electronic Engineering at University College London and Director of UCL's Adastral Park Postgraduate Campus. He is currently a holder of a Royal Society Wolfson Fellowship. He received his B.Sc. from University College London and Ph.D. from Oxford University. He was a member of the Technical Staff at AT&T Bell Labs at Murray Hill from 1984 until 1989 where his research interests were focused on mobile robots. In 1989 he joined NEC Research Institute in Princeton, NJ as a senior research scientist in the computer science division. At NEC, his research shifted to problems in computer vision and he was responsible for creating the computer vision group at NECI. He has worked on problems to do with stereo and motion correspondence and multimedia issues of image database retrieval and watermarking. In 1999, he was awarded the IEEE Signal Processing Society Best Paper Award (Image and Multidimensional Signal Processing Area) for a paper he co-authored on watermarking. From 1997-1999, he served as Chief Technical Officer of Signafy, Inc, a subsidiary of NEC responsible for the commercialization of watermarking. Between 1996 and 1999, he led the design of NEC's watermarking proposal for DVD video disks and later collaborated with IBM in developing the technology behind the joint "Galaxy" proposal supported by Hitachi, IBM, NEC, Pioneer and Sony. In 1999, he returned to NEC Research Institute as a Research Fellow. He is a Fellow of the IEEE, the IET (formerly IEE), and the British Computer Society. He is a member of the UK Computing Research Committee. He was founding co-editor in chief of the IEE Proc. on Information Security and is an associate editor of the IEEE Trans. on Information Forensics and Security. He is co-author of a book entitled "Digital Watermarking" and its second edition "Digital Watermarking and Steganography", and the co-editor of two books, 'Autonomous Robots Vehicles' and 'Partitioning Data Sets: With Applications to Psychology, Computer Vision and Target Tracking'.

IS-29
RETHINKING DIGITAL SIGNATURES

Ueli Maurer
Department of Computer Science, ETH Zurich, CH-8092 Zurich, Switzerland
maurer@inf.ethz.ch

Keywords: Digital signatures, digital evidence, non-repudiation, public-key certificates, public-key infrastructure (PKI), revocation, time-stamping.

Abstract: Digital signatures are a core enabling technology for evidence management in the context of the automation and digitization of business and government processes. The main advantages over conventional signatures are that they are easy to transmit, archive, search, and verify, and, moreover, promise to achieve a high level of security. Nevertheless the initial expectations for the usefulness of digital signatures were too optimistic. This calls for a systematic treatment of digital signatures and more generally digital evidence. The goal of this talk is to provide a foundation for reasoning about digital evidence systems and legislation, thereby identifying the roles and limitations of digital evidence.

1 INTRODUCTION

This short paper accompanies the author’s talk at ICETE 2008, which is based on and extends the article (Maurer, 2004). This short note only summarizes a few points at a non-technical level. Interested readers should consult (Maurer, 2004) and possibly also (Maurer, 2003).

This work draws its motivation from the apparent discrepancy between the potential usefulness of digital signatures as evidence and the fact that their use in practical applications is still very limited. We believe that a primary reason for this discrepancy is the lack of understanding of the semantics of digital evidence like digital signatures, certificates, time-stamps, etc. Our goal is therefore to provide a foundation for reasoning about digital evidence systems and legislation and to identify the roles and limitations of digital evidence, in the apparently simple scenario where it should prove that an entity A agreed to a digital contract d.

The approach proposed in (Maurer, 2004) is in sharp contrast to the current general views documented in the technical literature and in digital signature legislation. We propose an entirely new view of the concepts of certification, time-stamping, revocation, and other trusted services, potentially leading to new and more sound business models for trusted services. Some of the perhaps provocative implications of our view are that certificates are generally irrelevant as evidence in a dispute, that it is generally irrelevant when a signature was generated, that a commitment to be liable for digital evidence cannot meaningfully be revoked, and that there is no need for mutually trusted authorities like certification authorities.

We also propose a new type of digital evidence called digital declarations, based on a digital recording of a willful act indicating agreement to a document or contract.

2 EVIDENCE FOR NON-REPUDIATION

A basic act in business and other contexts is to enter a formal agreement, often called a contract, between two (or more) entities. Such an agreement requires the clear mutual understanding of all relevant parameters, in particular the terms and conditions. A contract is valid only if both parties formally entered it. It is generally understood that a contract has been entered by a user or entity only if he or she (or an authorized representative) performed a well-defined conscious and willful act, for instance by shaking hands and/or by signing a paper document, or by activating the generation of a digital signature on his smart-card.

In order to prepare for a possible future dispute, each party to a contract wants to keep sufficient evidence for the claim that the other party agreed to the contract. This is a symmetric goal, although the evidence collected by each party may be different. How
can an entity B obtain sufficient evidence that party A entered the contract? What sufficient means must be defined by the legal system.

One can distinguish at least three types of evidence, a combination of which may be used in a concrete setting: (1) physical evidence (e.g. a signed paper document), (2) statements by witnesses, and (3) digital evidence (digital signatures, time stamps, etc.). Physical evidence and witnesses seem to contradict the new paradigm of digitized business processes. A wide-spread expectation for digital signatures is that they allow to avoid the need for physical evidence and witnesses.

### 3 DIGITAL SIGNATURES: PROMISES AND OBSTACLES

A digital signature scheme works as follows. A user A generates a secret key, which she keeps secret, and the corresponding public key, which is made publicly available. The digital signature for a digital contract can be generated only when given the secret key, but it can be verified by anybody using the public key. Hence it can be interpreted as a proof that A agreed to the contract.

In view of the continuing automation and digitization of many business processes, the transmission, storage, and verification of physical evidence, like signed contracts, presents a major problem. In contrast to physical evidence, digital signatures are easy to transmit, archive, search, and verify. Moreover, digital signatures are generally unambiguous because their verification corresponds simply to the evaluation of a well-defined mathematical function, the signature verification predicate relative to a given public key. For these reasons, digital signatures promise to provide an elegant solution to the non-repudiation problem in the digitally operating economy.

Furthermore, due to the conjectured security of the underlying cryptographic mechanisms, digital signatures also promise substantially higher security compared to conventional signatures, and hence fewer disputes and simpler dispute resolution.

Despite the promises, in the context of non-repudiation services digital signatures are currently used only in isolated applications. We are still far from an internationally operational framework and infrastructure. Some of the obstacles are the lack of internationally applicable laws, the lack of standardization, the lack of viable business models for fostering the integration into business processes, and, last but not least, the abstractness and complexity of the subject matter, resulting in slow user acceptance.

An even more fundamental problem is that the meaning of digital evidence is not well understood. The intrinsic problem that a digital signature is not linked to any event in the real world has no solution. It is inherently impossible to determine when, where, how, and by whom a digital signature was generated, even when secure hardware, biometric identification, and time-stamping is used.

### 4 CONVENTIONAL VS. DIGITAL SIGNATURES

As a motivation for digital declarations, it is instructive to discuss the role of conventional hand-written signatures and why they are so useful in practice, despite the fact that their technical security is generally quite low. Conventional signatures are a pragmatic and flexible mechanism.

An idealized (but of course naïve) view of the use of conventional signatures can be described as follows. A user’s signature is well-defined, for instance by a master copy she has deposited. In case of a dispute, a signature allegedly issued by the user can be compared to her master signature. If one assumes that forged signatures can be recognized, then a signature is convincing evidence for the user’s consent to the signed document.

In practice, however, things are quite different from this idealized view. First, most people’s handwritten signatures are not very difficult to forge for a dedicated forger. Second, in most settings (except for example in a bilateral business relationship with a bank), a person’s master signature is neither deposited nor defined. Third, a person could use a signature different from the master copy in order to be able to later repudiate it. To avoid this problem, the receiving party would have to have on-line access to a master signature registry.

The value and purpose of a hand-written signature is not primarily that it is difficult to forge, but rather that it creates a situation in which a person A knows whether or not she signed, thus guaranteeing her awareness of performing a conscious and willful act. Similarly, forging a signature also requires a conscious act. Due to this guaranteed awareness, the denial of having signed a document is a precise and meaningful claim, equivalent to the (serious) claim that the signature is forged.

The described view on conventional signatures is in sharp contrast to what digital signatures can achieve. The existence of a digital signature does not imply the guaranteed awareness of the alleged signer
of the act that caused the signature generation. A signature could have been computed by a virus, because of another security problem, a flaw or ambiguity in the user interface, a flaw in the cryptographic mechanism, fraud or errors in the certification process, or any other of many possible reasons. Therefore a user A cannot meaningfully deny that a signature was generated by her. Rather, a denial is equivalent to the (for a judge quite useless) claim that she is not aware of having issued a signature.

5 DIGITAL DECLARATIONS

The dilemma can be described as follows. On one hand, the evidence collected to prove the execution of a contract should be digital in order to fit smoothly into the digitized business processes, but on the other hand it should be linked to the physical reality, in particular to a user’s willful act when agreeing to \( d \). This appears to be a contradiction, but there exists a solution proposed in (Maurer, 2003) and called digital declarations.

The user performs some willful act related to the relevant contract or document, and this act is recorded digitally and combined with characteristic information of the digital document. In a typical implementation, the digital declaration can be signed together with the actual digital document.

Digital declarations are purely digital and are hence easy to transport and store. But they offer the same functionality as conventional hand-written signatures. In particular, a user is guaranteed to be aware of what he is doing and can hence meaningfully be forced to deny a digital declaration. Such a denial is equivalent to the accusation that the digital declaration is forged.

Digital declarations can be embodied in many different ways. As an example, a user ordering a service or product on-line might be asked to speak a certain sentence referring to the product, the price, and the date of purchase. As another possibility, the willful act could be documented by a digital image, a video, or by any other recording device possibly invented in the future.

Digital declarations can be a useful feature of future digital transaction systems, for the following reasons:

- guaranteed user awareness,
- higher deterrence of misbehavior, hence fewer disputes,
- improved security compared to conventional signatures,
- lower cost due to reduced security requirements,
- improved acceptance of digital signature technology, and
- usability by moderately educated people.

REFERENCES


BRIEF BIOGRAPHY

Ueli Maurer is professor of computer science and head of the Information Security and Cryptography Research Group at the Swiss Federal Institute of Technology (ETH), Zurich. His research interests include information security, theory of cryptography (new paradigms, security proofs), applications of cryptography (e.g., digital signatures, public-key infrastructures, digital payment systems, e-voting), theoretical computer science, discrete mathematics, and information theory. He is also interested in the impact of IT on the society and economy at large. Maurer graduated in electrical engineering (1985) and received his Ph.D. degree in Technical Sciences (1990) from ETH Zurich. From 1990 to 1991 he was DIMACS research fellow at the Department of Computer Science at Princeton University, and in 1992 he joined the CS Department at ETH Zurich where he is a full professor. He has served extensively as an editor and a member of program committees. Currently he is Editor-in-Chief of the Journal of Cryptology, Editor-in-Chief of Springer Verlag’s book series in Information Security and Cryptography, and serves on the Board of Directors of the International Association for Cryptologic Research (IACR). He is a Fellow of the IEEE, a Fellow of the IACR, and was the 2000 Rademacher Lecturer of the Department of Mathematics at the University of Pennsylvania. Maurer has served as a consultant for many companies and government organisations, both at the management and the technical level. He serves on a few boards, including the board of directors of Tamedia, a Swiss media company, and the scientific advisory board of PricewaterhouseCoopers. He is a co-founder of the Zurich-based security software company Seclutions and holds several patents for cryptographic systems.
THE IMPORTANCE OF METROLOGY IN WIRELESS COMMUNICATION SYSTEMS

From AM/FM to SDR Systems

Nuno Borges Carvalho
Instituto de Telecomunicacoes / Universidade de Aveiro, Portugal

Abstract: Metrology is a key point in the design and optimization of wireless communication networks, in this talk we will make a brief presentation of the metrology chain necessities from base-band to RF wireless communication scenarios. In this path we will address linear and nonlinear measurements strategies, from RF figures of merit to base band figures. Finally some recent developments referred to software radio activities will also be presented, explaining the new paradigms and envisioning the future of wireless metrology.

BRIEF BIOGRAPHY

Nuno Borges Carvalho, was born in Luanda in 1972. He received the diploma and doctoral degrees in Electronics and Telecommunications Engineering from the Universidade de Aveiro, Aveiro, Portugal in 1995 and 2000 respectively. From 1997 to 2000 he was an Assistant Lecturer at the same University and a Professor since 2000. Currently he is an Associate Professor at the same University, and the scientific area coordinator of the wireless communications at Instituto de Telecomunicações. He has worked as a scientist researcher at the Instituto de Telecomunicações, and was engaged to different projects on nonlinear CAD, circuits and systems design and RF system integration. His main research interests include CAD for nonlinear circuits/systems, nonlinear distortion analysis in microwave/wireless circuits and systems and measurement of nonlinear phenomena, recently he has also been involved in design of dedicated radios and systems for newly emerging wireless technologies. In 2007 he was a visitor researcher at the North Carolina State University, and at the National Institute of Standards and Technology, NIST. Prof. Borges Carvalho is a member of the Portuguese Engineering Association and an IEEE Senior Member. He was the recipient of the 1995 University of Aveiro and the Portuguese Engineering Association Prize for the best 1995 student at the Universidade de Aveiro, the 1998 Student Paper Competition (third place) presented at the IEEE International Microwave Symposium, the 2000 IEE Measurement Prize. He is also the co-inventor of three registered national patents. He is a reviewer for several magazines including the IEEE Transactions on Microwave Theory and Techniques, IEEE Transactions on Circuits and Systems, IEEE Transactions on Communications and IEEE Transactions on Intelligent Transportation Systems, as some International journals from the Institute of Physics. He is the webmaster and an active member of the IEEE MTT-11 Technical Committee, and the chair of the URSI-Portugal metrology committee. Dr. Borges Carvalho is co-author of the book “Intermodulation in Microwave and Wireless Circuits” from Artech House, 2003.
Abstract: Future wireless communication systems should be operating mainly, if not completely, for burst data services carrying multimedia traffics. The need to support high-speed burst traffic has already posed a great challenge to all currently available air-link technologies based on either TDMA or CDMA. The current CDMA technology has been widely used in both 2G and 3G mobile cellular standards and it has been suggested that it is not suitable for high-speed burst-type traffic. There are many problems with current CDMA technology, such as its low spreading efficiency, interference-limited capacity and the need for precision power control, etc. This talk will address various important issues about the next generation CDMA technologies as a major air-link technology for futuristic wireless applications. In particular, it will cover two major topics: why we need the next generation CDMA technologies, and what is the next generation CDMA technology.

BRIEF BIOGRAPHY

Hsiao-Hwa Chen is currently a full Professor and was the founding Director of the Institute of Communications Engineering of the National Sun Yat-Sen University, Taiwan. He received BSc and MSc degrees from Zhejiang University, China, and PhD degree from University of Oulu, Finland, in 1982, 1985 and 1990, respectively, all in Electrical Engineering. He has authored or co-authored over 200 technical papers in major international journals and conferences, five books and several book chapters in the areas of communications, including the books titled "Next Generation Wireless Systems and Networks" (512 pages) and “The Next Generation CDMA Technologies” (468 pages), both published by John Wiley and Sons in 2005 and 2007, respectively. He has been an active volunteer for IEEE various technical activities for over 20 years. Currently, he is serving as the Chair of IEEE Communications Society Radio Communications Committee. He served or is serving as symposium chair/co-chair of many major IEEE conferences, including VTC, ICC, Globecom and WCNC, etc. He served or is serving as Associate Editor or/and Guest Editor of numerous important technical journals in communications. He is serving as the Editor (Asia and Pacific) for Wiley's Wireless Communications and Mobile Computing (WCMC) Journal and Wiley's International Journal of Communication Systems, etc. He is the founding Editor-in-Chief of Wiley’s Security and Communication Networks journal (www.interscience.wiley.com/journal/security). He is also an adjunct Professor of Zhejiang University, China, and Shanghai Jiao Tung University, China.
PATENTABILITY OF E-BUSINESS AND COMPUTER-IMPLEMENTED INVENTIONS AT THE EPO

Falk Giemsa
European Patent Office, Germany

Abstract: The majority of patent applications in the field of e-Business are refused at the European Patent Office (EPO). The reasons for the high number of refusals can be found in the European Patent Convention (EPC), in the decisions of the Boards of Appeal of the EPO and in the patent applications. This tutorial provides an overview of the general patent procedure before the EPO and explains the current work practice at the EPO for examining patent applications in the field of e-Business and Computer Implemented Inventions (CII).

BRIEF BIOGRAPHY

Falk Giemsa is a patent examiner at the European Patent Office in Munich. He studied computer sciences and now works in the field of data processing systems or methods, specially adapted for administrative, commercial, financial, managerial, supervisory or forecasting purposes and in the field of educational or demonstration appliances.
COMMUNICATION AND SOFTWARE TECHNOLOGIES AND ARCHITECTURES
FULL PAPERS
Keywords: Low-energy networking, ultra thin client, simulation.

Abstract: Power consumption of individual devices is often ignored in the development of computer networks. The traditional approach to a local area network, such as might be deployed in a cyber-café or classroom environment, has a number of workstations attached to a server, where the workstations have very similar specifications (and hence power demands) to the server. Intuitively, this means there is significant over specification of the workstation, and that much of the capability of the device is not used, with consequential wasted energy. Alternative topologies exist, particularly those which make use of thin client technology, and are meeting with success particularly in developing countries, where cost and power consumption take on a much greater importance. One such design is the ndiyo project, which is delivering thin-client based network solutions to a number of nations, allowing IT support to be deployed in places where it would not have otherwise been achievable. In this paper, we report our work in developing a simulation model to allow us to study the behaviour and operation of ndiyo, offering us the opportunity to carry out some “what if” analyses of the behaviour of such systems under differing network loads.

1 INTRODUCTION

The typical workstation / server approach to the provision of networked IT, in which a number of “workstations”, each of which is actually a stand-alone computer system, are networked to a “server” which acts as a central repository of files and as a gateway to the wider network (the Internet) is very well established. However, this design, whatever its advantages, appears inefficient from the perspective of resource utilisation. In particular, in situations such as the typical cyber café, and also in classroom situations where web browsing is the prevalent activity, workstations are effectively little more than smart terminals. The use as workstations of devices similar in power to the server, suggests that significant resource capacity may not be being used. While this may be acceptable from the perspective of resource use, since the machines are relatively easily available; from the viewpoint of energy use, different arguments can be made. In situations where energy and other resources are at a premium, it makes sense to seek out other mechanisms to deliver the required processing power.

One alternative to the widespread deployment of full-specification systems uses thin client technology, taking advantage of the fact that most of the processing can be carried out by the server, so the client’s power can be reduced accordingly.

2 THIN CLIENT NETWORKS

2.1 Introduction

Thin client Networks have a long history; the development of terminal servers is one area which has a long association with thin client networks.

A Thin client (sometimes also called a lean client) is that part of client-server architecture networks which depends primarily on the central server for processing activities, and mainly focuses on conveying input and output between the user and the remote server. In contrast, a thick or fat client does as much processing as possible and passes only data for communications and storage to the server.

Many thin client devices run only web browsers or remote desktop software, meaning that all significant processing occurs on the server.

However, recent devices marketed as thin clients can run complete operating systems such as Debian..
GNU/Linux, qualifying them as diskless nodes or hybrid clients (Wikipedia, 2007). The emphasis of this paper is on such thin client networks, in fact the devices discussed and tested in this paper are referred to as Ultra-thin client devices. (Nidyo, 2006)

2.2 Related Work

Previous work in the field has addressed a variety of topics which have direct impact upon the operation of thin client systems. Kelly (2002) reports on the need to gather appropriate behavioural data to use as input into a simulation system. Tolia et. al. (2006) make reference to the fact that “adequacy of thin-client computing is highly variable and depends on both the application and the available network quality”. The increasing deployment of wireless and hand-held (battery powered) devices has led to consideration of whether thin-client technology is appropriate for such technologies, and measurements have been conducted by Yang et. al. (2003) reporting that such systems can operate successfully even with the relatively high packet loss rates which can be experienced with wireless networks, and by Lai et. al, (2004) reporting lower bandwidth requirements and hence better user experience for thin client systems.

In light of these and other papers, we determined to conduct a simulation-based experiment using a particular thin-client implementation, in part to explore the performance of such a thin-client system, but also to determine the adequacy of simulation methods in this application.

3 NDIYO PROJECT

This paper is focused on Nivo devices which are ultra thin client devices developed for the Ndiyo Project by Displaylink. Ndiyo is a not for profit Cambridge based project which aims to provide affordable and sustainable IT networks to the world. The following section is an extract from the ndiyo website (Ndiyo, 2006).

3.1 “Nivo” Ultra Thin Client

A device called a Nivo (Network In, Video Out) is a highly optimized piece of electronics, dedicated to the purpose of displaying an interactive computer desktop over a network. The server simply sends to the nivo - over the network using a simple compression scheme - the pixels that need to be displayed on the user’s screen. Modern wired Ethernet networks (100Mbit or higher) are fast enough for this approach to keep the user's screen up to date. Nivo simply has an Ethernet (network) socket, a low voltage power socket, sockets for keyboard and mouse, and a VGA (monitor) socket. This box replaces the entire PC in a conventional computer workstation (Ndiyo, 2006).

3.2 The Network

We were given the task to test the performance of a nivo network with between 2 to 30 nivo devices attached to a single server. Below are the major aspects of the nivo device’s network topology and their operations:

- The Nivo network generally runs on a switched 100 Mb/s LAN, as Nivo devices usually support 100 Mb/s.
- In Nivo networks the transmitting packets have areas of pixels, compressed in a lossless way, transmitted over a simple transport layer which provides very basic reliability - much simpler than TCP. An approach of VNC transmitted over UDP is best approximate for this situation.
- An update sent out from the server can be as large as a whole screen refresh – perhaps 5MB – or as small as a few bytes. The big updates will be split into chunks of approximately1500 bytes when carried in an Ethernet UDP packet. Keystroke and mouse events coming back are usually small, mouse event reports are around 10 bytes and reported at most about 50 times per second.

3.3 Prerequisite

The major focus of this study was to predict the real network bandwidth usage or at least some reasonable approximations. However the aim of the study is not only the network bandwidth usage but other characteristics as well, in addition we wish to measure those characteristics in different scenarios. Some recommended scenarios are:

1. Large number of users using voice conversation and web applications.
2. Large number of users using video and web applications.
3. Large number of users using only web applications.
4 SIMULATION MODEL

On the basis of the above defined structure of nivo networks we have developed a simulation model in to predict the usage of network bandwidth in different scenarios. The details of the simulator and the model developed are discussed further.

4.1 NS-2

All the simulations in this project are developed with the help of Network Simulator 2 (NS-2). The latest version of NS-2 i.e. NS-2.30 was used.

NS-2 is an event driven network simulator developed by UC Berkeley. NS-2 implements traffic behaviours, network protocols, routing, etc. for simulation. Because it is open source software, during the development many contributions have been included from other researchers. NS-2 has become a common tool for network researchers to simulate and evaluate network related project.

Through the OTcl language interface users can define a particular network topology, the protocols and applications that they want to simulate and the form of the output that they want to obtain from the simulator quickly and clearly as a script. (Zhao and Wu, 2006).

4.2 Thin Client Simulation Model

The available information about the network clearly states that the Server should be sending a reasonable amount of data at different intervals towards its client on their requests, so the transmission of data is duplex and in intervals from both sides.

The simulation model was designed in two stages. First the topology of the network was developed and tested (Figure 1). It is obvious from the diagram that the topology here is that of a switched LAN. The ideal bandwidth for an Ethernet LAN is 100 Mb/s but generally performance is nearer to 85 Mb/s due to different losses. After selecting User datagram Protocol (UDP) packets as the major traffic on the network some suitable traffic generators were attached to both entities i.e. The Server and the Workstations. As the data generated from both sides is in intervals a Pareto traffic generator is used, generating traffic using a probability density function i.e.

$$Pr(X > x) = \left( \frac{x}{x_m} \right)^{-k}$$

This is sufficient to generate traffic at random intervals but can be associated with other intervals by its time interval parameters. The packet sizes are designated according to the given information i.e. 1500 bytes size generated from the server and 10 bytes size generated from the workstations. The rate at which these packets are generated depends upon applications used by the users at the server.
5 OPERATION

5.1 Overview

In order to describe the operation reference is again made to Figure 1. As soon as a workstation sends a request to the server while running or initiating an application the server starts responding by sending a large amount of data in chunks of 1500 bytes with assigned intervals i.e. with a delay of 200ms or according to the application. There is a separate traffic generator for each user therefore the volume of data generated increases with the number of users. Another major aspect in Traffic Generation is the profile with which the generator is generating data. This option is easily available in Pareto & CBR traffic generators within NS-2.

The workstations are connected with the server through an Ethernet switch hence the network monitoring is performed at this point. In order to do so we have used a perl script, this script calculates the throughput at our preferred node at given time intervals. The node selected here as mentioned earlier is our network switch because all of the network data will be flowing through it and the time intervals of granularity at which the throughput of switch was calculated is 1 sec.

5.2 Test Bed

Before developing our simulation according to the Ndiyo scenario we created a test bed within our lab. We wanted to compare our simulation data rates along with the data rates of that traffic which was generated while using VNC (Virtual Network Computing) software. This utility is widely used in order to achieve remote desktop facility over different platforms i.e. it is platform independent. The operation of VNC is almost the same as that of the software used by the Ndiyo project, although VNC provides a sophisticated connection setup and delivery of data while using TCP. This differs from the Ndiyo project which uses a VNC over UDP approach in their software development. However, for the purpose of exploring the activity of VNC, this variation is not important, since we are not measuring the relative performance or reliability.

We used Fedora Core4 and VNC packages downloaded from www.Realvnc.com. VNC provides a VNCVIEWER used at the user side and the VNCSERVER to be run at the server. VNCVIEWER provides an X11 session transmitted from the server.

Different sessions of testing within our lab provided us some more useful information such as if the server is not able to support a large amount of multiprocessing, this creates delay and a bottleneck type situation. However this is not caused by the lack of network bandwidth but to the low power of the server (a Pentium4). In order to monitor traffic Ethereal was used at the server but we were able to monitor only that traffic associated to the server and connected users.

6 RESULTS

The data rates were recorded during our traffic generation, which was mainly HTTP traffic, but including some web streaming as well. These data rates were used while producing the simulation according to our test bed as the Lab Network is almost same as our simulation model. Figure 2 shows that there are packet losses ranging from 2 to almost 3 Mbits in short time intervals. The highest peak achieved is just more than 3.5 Mbits which is well within the actual network bandwidth of around 85Mb (eliminating different network losses from a typical 100Mb LAN), but the processing of the systems is taken error free as a default in these simulations. With little difference the results from the simulation compare with our measured outcomes.

When the same data rate was used for 30 users in our simulation model the output that was achieved is depicted in Figure 3. The throughput was almost ten times the 2 user throughput but yet again never exceeded any problematic threshold. Another interesting aspect of Figure 3 is that there were no major packet or throughput drops during the simulation. We believe that the small number of applications used in our test bed meant the data rates were at an average level.

As mentioned earlier, our simulations were developed using data rates procured from our test bed and with minimal applications used in order to extend the data patterns using run length encoding to create our own scenarios. Doing this, we developed a simulation which produced a throughput as depicted in Figure 3. These steps were taken in order to create simulations which were directly related to a practical scenario.

However in order to evaluate the bottleneck situations it was necessary to create some hypothetical scenarios where a number of applications are used with each one having high data generation rates. At this point it should be
understood that as NIVO devices are in their test phase and are not produced in large quantities, we were unable to get our hands on NIVOs. To collect such of data rates we used another network monitoring tool OBSERVER, which provided us the facility to record data rates for different applications while using Remote Desktop Control. Data rates for different applications are shown in Table 1; however as these data rates were collected while using Remote Desktop Control the amount of TCP packets is also included.

Table 1: Data rates used.

<table>
<thead>
<tr>
<th>Remote Video file</th>
<th>Remote Web Streaming</th>
<th>Remote HTTP Traffic</th>
</tr>
</thead>
<tbody>
<tr>
<td>500-2450 Kb/s (25 fps)</td>
<td>727-900 Kb/s</td>
<td>97-150 Kb/s</td>
</tr>
</tbody>
</table>

These rates were obtained while using a single user, in order to create scenarios for large number of user the traffic generator we can easily run-length-encode these rates. Also the data generation rates depend on the nature of the file or web page, if the data set is too rich then more pixel data has to be sent over the network and this could cause an increase in the data rates. Therefore different type of video files, streams or web pages could generate different amount of data. Hence the above examples can be best described as samples for such type of traffic. To create such a scenario where users are using different applications at the same time, in order to evaluate when a bottleneck situation occurs, we developed simulations with combination of the above data rates.

Figure 4 depicts a condition where only 2 users are using the network. The amount of throughput is almost double that recorded in previous simulations. Whereas Figure 5 shows the situation where 30 users at a time are connected to the network and utilizing different services.

It can be easily noticed that due to the increased volume of data there is a significant amount of packet drop; Figure 5 shows that there are packet drops from around 6 Mbits to more than 10 Mbits with in a very small period of time. However, it is obvious that with a small number of users, the available bandwidth is sufficient to support all of them. The packet drops which are noticed at different intervals during the simulation mainly affect video file transmission, and could result in jerks in a multimedia file. These packet drops can also be a result of reduced processing capability at the server as the number of applications used increases.

7 CONCLUSIONS

A comparison of the four graphs reveals some very obvious but important points:

- A LAN network can easily deliver multimedia services to its users, but in case of Nivo devices all of the processing is carried at the server end and therefore data sent to the user from the server is larger. Clearly, in the case of a video file being executed by a user while using network neighborhood facility all the processing is done at the user end while only the file data is being taken from the remote location, whereas in our scenarios the whole processed screen is sent over the network which ultimately results in large amount of data transmission over the network.

- Relating to our previous argument, if we evaluate our thin client networks then it is obvious that large number of user could be accommodated by the network but only if we execute those applications which generate lower data rates. Multimedia applications can only be executed if there is a small number of users (5-10). Referring to our test bed results, web streaming can be achieved but not if a large number of users are using it simultaneously.

- Thin client networks such as Ndiyo can provide affordable network solutions to those areas where a user’s requirements are more related to HTTP traffic rather than multimedia applications.

REFERENCES


The Ndiyo Project. 2006 http://www.ndiyo.org/systems
<last accessed 08/04/08>
Tolia, N. Andersen, D. Satyanarayanan 2006. Qualifying
Interactive User Experience on Thin Clients IEEE
Wikipedia definition of Thin client. From Wikipedia,
accessed 08/04/08>

Yang, S.J., Nieh J., Krishnappa, S., Mohla, A., Sajjadpour,
M. 2003, Web Browsing Performance of Wireless
Thin-Client Computing Proceedings of 12th
International World Wide Web Conference Budapest,
pp. 68-79.

Figure 2: Data for two simultaneous streaming users.

Figure 3: Data for 30 simultaneous streaming users.
Figure 4: Data for two simultaneous video users.

Figure 5: Data for 30 simultaneous video users.
USING CO-OCCURRENCE TO CLASSIFY UNSTRUCTURED DATA IN TELECOMMUNICATION SERVICES

Motoi Iwashita, Ken Nishimatsu and Shinsuke Shimogawa
NTT Service Integration Laboratories, Midori-cho 3-9-11 Musashino, Tokyo, Japan
iwashita.motoi@lab.ntt.co.jp, nishimatsu.ken@lab.ntt.co.jp, shimogawa.shinsuke@lab.ntt.co.jp

Keywords: Co-occurrence, Correspondence analysis, Telecom operation, Text mining.

Abstract: A variety of services have recently been provided according to the highly-developed networks and personal equipment. Connecting this equipment becomes more complicated with advancement of these day by day. Because software is often updated to keep up with advancements in services or security, problems such as no-connection increase and determining the cause become difficult in some cases. Telecom operators must understand the situation and act as quickly as possible when they receive customer enquiries. In this paper, we propose one method for analyzing and classifying customer enquiries that enables quick and efficient responses. Because customer enquiries are generally stored as unstructured textual data, this method is based upon a co-occurrence technique to enable classification of a large amount of unstructured data into patterns.

1 INTRODUCTION

A conventional fixed telephone service is simply provided by a telephone network. Because the network structure is simple, it is easy to determine the cause of service problems. Furthermore, telecom operators with long-accumulated know-how can act quickly. Recently, broadband infrastructures with Asymmetric Digital Subscriber Line (ADSL) and optical fibers have penetrated the telecommunications industry. This trend has induced the expansion of a variety of services, such as the exponential use of the Internet, the provision of Voice over Internet Protocol (VoIP) and video distribution services, and security software countermeasures against virus attacks on PCs. Therefore, the end-to-end network structure has become complicated, considering the connection of home equipment, such as a modem, and the setup of its software. As a result, discovering the cause of problems is difficult. Connecting service equipment will continue to become more complicated in the near future, if we consider the drive forwards ubiquitous services.

Customer satisfaction decreases when a long time is spent on restoration because discovering the cause is difficult. Therefore, it is necessary to understand the features of the problem first to resolve the situation by data classification.

The saved information is not classified as structured data, i.e., it is an unstructured data. Let us take a customer phone enquiry about no-connection as an example. There are several causes, such as failure of the optical fiber, modem, or application. Both normal and strange situations are included in text such as the connection to the Internet is OK, but e-mail cannot be sent. Therefore, it is important to classify the unstructured textual data with accuracy.

A text mining technique, such as morphological analysis, syntax analysis, co-occurrence relation, etc., is effective (Ohsumi, 2006), (Sato et al., 2007), (Sullivan, 2001), (Toda et al., 2005). This technique is applicable to customer questionnaire analyses in product development, word searches in portal sites such as Google and Yahoo, term frequency analyses in web logs (blogs) or customer generated media, article classification by keyword in news articles, and evaluation indexes of a company’s image. Mainly morphological analysis is applied in these areas to survey trends by analyzing the frequency of terms in selected text. A keyword is extracted as a topic of a sentence in terms of the features of the network structure (Masuo et al., 2001), (Ohsawa et al., 1997), (Cutting et al., 1992), (Ho et al., 2001), (Leuski, 2001). Clustering and co-occurrence related methods have been proposed classify keywords and relate them to synonymous terms, different words having the same meaning, and synonyms, which have similar meanings (Uejima et al., 2004), (Rodriguez et al., 1998).
An improved method was proposed for synonymous term classification in fuzzy searches for the aim of failure analysis (Naganuma et al., 2005). Simply understanding failure trends and noting customer requirements when analyzing an enquiry then analyzing the word trends is not always effective. Understanding the meaning of sentences is essential. There are no effective methods to semantically analyze text that are applicable to telecom management.

A text classification method from a semantic point of view that considers the features of telecom services and the co-occurrence of terms for classifying and analyzing a large amount of unstructured data consisting of customer enquiries is proposed. The difficulties in analyzing textual data in telecom services when conventional techniques are used is explained in Sec. 2. Section 3 describes the features of telecom services. Our classification method is proposed in Sec. 4, and the results are discussed in Sec. 5.

2 DATA ANALYSIS IN TELECOM OPERATIONS

2.1 Necessities of Data Classification and Text Mining

In general, the telecom operator saves the customer enquiry and the coping process as information. The aim of this is to enable the finding of similar problems by searching with related keywords and to enable quick action when such problems occur. These coping processes are effective for sharing the knowledge among assigned telecom operators and for improving their skills. Therefore, this method is useful when problems happen. However, the drawback of this method is that it is impossible to get an overview of all the possible patterns of a problem and to establish coping processes for more complex problems in advance. Therefore, it is necessary to establish an effective coping process for customer enquiries, assign optimal operators due to advance classification of customer enquiries and survey the failure trend.

Information generally consists of text. It takes a long time to analyze text word by word and to classify large amounts of textual data. Therefore, an effective method based on a text mining technique, such as term frequency analysis, number of synonymous words determination and related terms extraction is needed.

2.2 Limitations of Morphological Analysis

Figure 1 shows the relationship between term frequency and its ranking for 10,000 customer enquiries about telecom services. The terms were classified and counted by morphological analysis, and those that appear more than 50 times are shown in Fig. 1.

Figure 1: Relationship between term frequency and its ranking.

"Category 1" is text relating to equipment, such as a modem, PC, telephone, etc., whereas "Category 2" is text relating to a failure phenomenon, such as setup trouble, no-connection, cable breakdown, etc. "Category 0" means terms analyzed without considering categories.

All the patterns show a \( n/1 \) feature, so they follow a power law (Newman, 2005), (Zipf, 1949), as is normal in general sentences. A power law implies that there are many kinds of terms in the textual data and means that there are many types of customer enquiries. Since the slope of Category 1 is steeper than the other two, the terms used in that textual data are limited compared with the terms in the other categories. On the other hand, the slope of Category 2 is gradual, which shows the variety of failure phenomenon terms. These results show that it is possible to only survey the terms with high frequency, such as PC, telephone, internet, etc. However, it is necessary to understand the relationships among the terms.

2.3 Limitations of Correspondence Analysis

To classify into groups by term features, correspondence analysis (Benzecri, 1992), (Hayashi, 1993), (Takahashi, 1996) is applied. The terms with frequency rankings higher than 400 were selected from
10,000 customer enquiries. The analysis of up to the 10th factor is shown in Table 1, and a graph of the 1st and 3rd factors is shown in Fig. 2.

i-th contribution rate, \( r_i \), is calculated as follows.

\[
 r_i = \frac{(\text{Eigenvalue}(i))}{\sum_{j}(\text{Eigenvalue}(j))} \tag{1}
\]

These results show the difficulty of describing the features. This is because the accumulated contribution rate is under 20%, even considering the 10th factor, and most of the information is gathered in a small area. Therefore, it is clear that we need some preparation before classification.

Table 1: Correspondence analysis.

<table>
<thead>
<tr>
<th>Factor</th>
<th>Eigenvalue</th>
<th>Contribution rate</th>
<th>Accumulated rate</th>
</tr>
</thead>
<tbody>
<tr>
<td>1st</td>
<td>0.7466</td>
<td>2.36%</td>
<td>2.36%</td>
</tr>
<tr>
<td>2nd</td>
<td>0.6999</td>
<td>2.12%</td>
<td>4.47%</td>
</tr>
<tr>
<td>3rd</td>
<td>0.6181</td>
<td>1.95%</td>
<td>6.43%</td>
</tr>
<tr>
<td>4th</td>
<td>0.5914</td>
<td>1.87%</td>
<td>8.30%</td>
</tr>
<tr>
<td>5th</td>
<td>0.5903</td>
<td>1.83%</td>
<td>10.13%</td>
</tr>
<tr>
<td>6th</td>
<td>0.5561</td>
<td>1.77%</td>
<td>11.90%</td>
</tr>
<tr>
<td>7th</td>
<td>0.5469</td>
<td>1.73%</td>
<td>13.62%</td>
</tr>
<tr>
<td>8th</td>
<td>0.5279</td>
<td>1.67%</td>
<td>15.29%</td>
</tr>
<tr>
<td>9th</td>
<td>0.5301</td>
<td>1.64%</td>
<td>16.93%</td>
</tr>
<tr>
<td>10th</td>
<td>0.5083</td>
<td>1.61%</td>
<td>18.54%</td>
</tr>
</tbody>
</table>

Figure 2: Relationship between 1st and 3rd factors.

3 FEATURES OF TEXTUAL DATA IN TELECOM SERVICES

The features of textual data in telecom services are explained in this section. An operator writes down information about a customer enquiry. Therefore, the style of the description deeply depends on the operator. That is, the description style is not ordered. There may be abbreviations of terms, wording that only the specified operator can understand, and so on. Moreover, there are many synonymous terms. Taking service specifications of optical fibers as an example, there are descriptions using general names (e.g. optical service, Fiber-To-The-Home (FTTH)), and special/abbreviated names for customer enquiries (e.g. description of product name, abbreviation). Let us consider a situation where optical and telephone services are provided to a customer. It is difficult to determine whether the cause of a problem is an optical cable, modem failure, or application problem in the case of a phone call from a customer about no-connection. Moreover, associated factors that complicate the situation include partial trouble, e.g. the Internet works whereas e-mail does not, or no-connection because of software compatibility.

To summarize, it is difficult to apply a text mining technique directly to raw textual data in telecom management for semantic/structural classification. Therefore, we need a modification to distinguish the features of telecommunication services. Telecommunication service such as internet connection, VoIP, is generally provided by an end-to-end network consisting of a telephone, PC, the carrier’s network, the provider’s server, etc. Fig. 3. There is clearly an event feature for each component of the network. We can predict that the component, such as the service, telephone, PC, and network, is strongly related to the problem, such as failure, misconfigured of set-up, and cable breakdown, respectively. Therefore, by designating the network factor as one event and the problem as the other event, we can construct a semantic representation. Moreover, if a problem occurs in one piece of equipment in a network, it is expected to lead to problems in the other component or to other events.

Figure 3: Telecom features categorization.
4 CLASSIFICATION METHOD BASED ON CO-OCCURRENCE

4.1 Framework of Classification

The requirement for textual data classification is based on the ability to cover all textual data and to fit the operator’s thinking. It is desirable that this classification be determined from the viewpoints of term frequency, co-occurrence, and cause-effect relationship, as shown in Table 2. Term frequency can tell us what kind of customer enquiries often appear, while co-occurrence tells us which terms are strongly related. Cause-effect relationship tells us the relationship among multiple terms such as network component and problem event.

Table 2: Criteria for classification.

<table>
<thead>
<tr>
<th>Criteria</th>
<th>Feature</th>
</tr>
</thead>
<tbody>
<tr>
<td>Frequency of a pair of terms</td>
<td>Number of appearances in textual data</td>
</tr>
<tr>
<td>Co-occurrence of a pair of terms</td>
<td>Relationship of a pair of terms</td>
</tr>
<tr>
<td>Transition rate among multiple terms</td>
<td>Cause-effect relationship among terms</td>
</tr>
</tbody>
</table>

Figures 4 and 5 show the frameworks in which textual data is classified in terms of the previous three criteria.

Classification framework:

- Procedure 1: Classification by type of access (Fig. 4 (a))
  - Textual data is classified in terms of access, i.e., type of service, such as dial-up, ADSL, FTTH, and so on.

- Procedure 2: Classification by category (Fig. 4 (b))
  - Textual data is classified in terms of categories based on morphological analysis. For example, Category A means component, such as a service, telephone, PC, modem, etc., while Category B means the problem, such as no-connection, misconfiguration, etc.

- Procedure 3: Calculation of term frequency and co-occurrence (Fig. 5 (a))
  - The frequency of both $x$ and $y$ appearing in textual data is represented by $f(x,y)$. Let us select a pair in terms of a frequency greater than $\beta$, where $\beta$ is a given threshold. Then, calculate co-occurrence as follows:

$$C(x,y) = \frac{f(x,y)}{f(x) + f(y) - f(x,y)}$$  \hspace{1cm} (2)

for any $x$ and $y$ such that $x,y \in A$ or $B$.

- Procedure 4: Transition among multiple terms (Fig. 5 (b))
  - Let us choose a pair of terms such that $C(x,y) \geq \alpha$, where $\alpha$ is a given threshold.
  Step 1: Select a pair with Categories A and B satisfying $\alpha \leq C(i,j)$, $i \in A$, and $j \in B$.
  Step 2: Select a pair with Category A satisfying $\alpha \leq C(i,j)$, $i, j \in A$, and $\alpha \leq C(i,k)$, $i \in A$, and $k \in B$.
  Step 3: Select a pair with Category B satisfying $\alpha \leq C(i,j)$, $i, j \in B$, and $\alpha \leq C(i,k)$, $i \in B$, and $k \in A$.

Figures 4 and 5 show the frameworks in which textual data is classified in terms of the previous three criteria.

Classification framework:

- Procedure 1: Classification by type of access (Fig. 4 (a))
  - Textual data is classified in terms of access, i.e., type of service, such as dial-up, ADSL, FTTH, and so on.

- Procedure 2: Classification by category (Fig. 4 (b))
  - Textual data is classified in terms of categories based on morphological analysis. For example, Category A means component, such as a service, telephone, PC, modem, etc., while Category B means the problem, such as no-connection, misconfiguration, etc.

- Procedure 3: Calculation of term frequency and co-occurrence (Fig. 5 (a))
  - The frequency of both $x$ and $y$ appearing in textual data is represented by $f(x,y)$. Let us select a pair in terms of a frequency greater than $\beta$, where $\beta$ is a given threshold. Then, calculate co-occurrence as follows:

$$C(x,y) = \frac{f(x,y)}{f(x) + f(y) - f(x,y)}$$  \hspace{1cm} (2)

for any $x$ and $y$ such that $x,y \in A$ or $B$.

- Procedure 4: Transition among multiple terms (Fig. 5 (b))
  - Let us choose a pair of terms such that $C(x,y) \geq \alpha$, where $\alpha$ is a given threshold.
  Step 1: Select a pair with Categories A and B satisfying $\alpha \leq C(i,j)$, $i \in A$, and $j \in B$.
  Step 2: Select a pair with Category A satisfying $\alpha \leq C(i,j)$, $i, j \in A$, and $\alpha \leq C(i,k)$, $i \in A$, and $k \in B$.
  Step 3: Select a pair with Category B satisfying $\alpha \leq C(i,j)$, $i, j \in B$, and $\alpha \leq C(i,k)$, $i \in B$, and $k \in A$.

Figures 4 and 5 show the frameworks in which textual data is classified in terms of the previous three criteria.

Classification framework:

- Procedure 1: Classification by type of access (Fig. 4 (a))
  - Textual data is classified in terms of access, i.e., type of service, such as dial-up, ADSL, FTTH, and so on.

- Procedure 2: Classification by category (Fig. 4 (b))
  - Textual data is classified in terms of categories based on morphological analysis. For example, Category A means component, such as a service, telephone, PC, modem, etc., while Category B means the problem, such as no-connection, misconfiguration, etc.

- Procedure 3: Calculation of term frequency and co-occurrence (Fig. 5 (a))
  - The frequency of both $x$ and $y$ appearing in textual data is represented by $f(x,y)$. Let us select a pair in terms of a frequency greater than $\beta$, where $\beta$ is a given threshold. Then, calculate co-occurrence as follows:

$$C(x,y) = \frac{f(x,y)}{f(x) + f(y) - f(x,y)}$$  \hspace{1cm} (2)

for any $x$ and $y$ such that $x,y \in A$ or $B$.
4.2 Transition of Relativity Among Multiple Terms

The effectiveness of procedure 4 in the previous section is explained in this section. For example, a customer makes a general complaint that he/she is not able to send e-mail from his/her PC as there is no connection with the Internet. The cause might not be in the PC but in the modem setup in that case. Therefore as a transitional way of thinking, we need relative keywords to suggest other causes. The relationship among multiple terms should be clear. If we calculate all co-occurrence values between the pairs of terms in Categories A and B, we need a long calculation time $n^2 \times n^2 = o(n^4)$. If we use procedure 4, on the other hand, we can reduce the calculation time to $n \times n = o(n^2)$. This is because co-occurrence is calculated for each category as a unit.

5 DISCUSSION OF RESULTS

5.1 Co-occurrence in a Category

The order of choice strongly depends on the threshold in classification procedures. The relationship between term frequency and co-occurrence in textual data from 1000 customer enquiries is shown in Table 3. 1000 customer enquiries are the field data saved by telecom management for IP-based services during the daytime at one day in 2007. Pairs of terms are ordered by term frequency. There are three cases for co-occurrence with the threshold as a parameter. The selected pairs become similar in term frequency when the threshold decreases. Because we screened and chose pairs with high frequency by procedure 2, using the choice decided by co-occurrence was effective for representing the features of the text. The threshold is given high value at first step. If the number of pairs is small, $\alpha$ decreases as a second step. In this way, iteration step of decreasing $\alpha$ can lead relationship among terms.

<table>
<thead>
<tr>
<th>Pair of terms</th>
<th>Term frequency</th>
<th>Co-occurrence</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>$\alpha = 0.02$</td>
</tr>
<tr>
<td>1A&amp;2A</td>
<td>1st choice</td>
<td>2nd</td>
</tr>
<tr>
<td>1A&amp;4A</td>
<td>2nd</td>
<td>-</td>
</tr>
<tr>
<td>4A&amp;5A</td>
<td>3rd</td>
<td>1st</td>
</tr>
<tr>
<td>1A&amp;5A</td>
<td>4th</td>
<td>-</td>
</tr>
<tr>
<td>2A&amp;4A</td>
<td>5th</td>
<td>-</td>
</tr>
</tbody>
</table>

5.2 Co-occurrence Among Categories

Co-occurrence among categories was calculated for the textual data from 1000 customer enquiries, as shown in Fig. 6. We chose the pairs marked that have a value more than the given threshold. Then, seven pairs of terms were selected by co-occurrence. Moreover, there were pairs that had a strong transition rate in the same category. We merged those pairs and selected them as the 8th and 9th choices Fig. 6.

Let us compare the proposed method and the method using only term frequency for selecting pair of terms. The frequency of 8th choice is 70 by pair of terms-frequency, while that of the 8th choice by the proposed method is 54. Because the difference between two choices is small in that amount of data and the 8th choice by pair of terms-frequency is smaller than the given threshold, it only represents weak relationships in the features of the text. The proposed method is possible to classify and understand complicated structure through selecting and relating strong co-occurrence. Therefore, the choice by co-occurrence represents the features of the text.

<table>
<thead>
<tr>
<th>No.</th>
<th>Pair of terms</th>
</tr>
</thead>
<tbody>
<tr>
<td>1st choice</td>
<td>1A&amp;2B</td>
</tr>
<tr>
<td>2nd choice</td>
<td>4A&amp;1B</td>
</tr>
<tr>
<td>3rd choice</td>
<td>2A&amp;2B</td>
</tr>
<tr>
<td>4th choice</td>
<td>1A&amp;1B</td>
</tr>
<tr>
<td>5th choice</td>
<td>2A&amp;1B</td>
</tr>
<tr>
<td>6th choice</td>
<td>4A&amp;4B</td>
</tr>
<tr>
<td>7th choice</td>
<td>4A&amp;2B</td>
</tr>
<tr>
<td>8th choice</td>
<td>(1A and 2A)&amp;(1B and 2B)</td>
</tr>
<tr>
<td>9th choice</td>
<td>(1A and 4A)&amp;(1B and 2B)</td>
</tr>
</tbody>
</table>

Figure 6: Relationship between categories.
old as a parameter. The number of candidates increases when the threshold decreases. This is because of the weakness of co-occurrence. The number of pairs with a transition rate grows when the threshold decreases. The 6th choice is (1A and 2A) & (1B and 2B) when $\alpha = 0.25$. 1A and 2A correspond to “Internet” and “VoIP” respectively, while 1B and 2B correspond to "connection is OK" and "no connection" respectively. We can classify the text in a semantic sense, e.g., "Internet is OK, but VoIP has no connection".

This type of proposed data classification can get an overview of all the possible patterns of a problem and establish coping processes in advance. Furthermore, it has possibilities to mine the potential customer requirement that leads to new business.

![Figure 7: Threshold and pairs.](image)

### 6 CONCLUSIONS

A classification technique for customer enquiries is needed due to the increasing complexity of the connections in end-to-end networks in the telecom operating field. In this paper, we proposed one method for analyzing and classifying customer enquiries that enables quick and efficient responses. Because customer enquiries are generally stored as unstructured textual data, this method is based upon morphological analysis and co-occurrence techniques to enable classification of a large amount of unstructured data into patterns. We applied the proposed method to 1000 customer enquiries and evaluated its effectiveness. The method can apply not only to establish coping processes in advance but also to mine potential requirement for new business.

We are currently conducting further study on applying this method to large amounts of data and on determining a threshold for telecom operation.

### REFERENCES


SHORT PAPERS
FLEXIBLE DATA SEARCHS USING CONDITION FORMULAS

Toshio Kodama¹, Tosiyasu L. Kunii² and Yoichi Seki³

¹Venture Business Department, Maeda Corporation, 2-10-26 Fujimi, Chiyoda-ku, Tokyo, Japan
kdoama.ts@city.maeda.co.jp
²Morpho, Inc., The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo, Japan
tosi@kuni.info, tosi@kunii.com
³Software Consultant, 3-8-2 Sakaemachi, Hino City, Tokyo, Japan
yseki@cbl@mail.hinocatv.ne.jp

Keywords: Cellular model, formula expression, topological space, cellular space, condition formula.

Abstract: Cyberworlds are distributed systems on the Web, and are constantly evolving like living things, creating value. Currently, numerous Web business applications, such as cyberworld systems are being built, but in the development of the systems, combinatorial explosion happens because schemas and application programs must be modified whenever schemas change. We designed and implemented the logic of a flexible data search function by employing a condition formula on the cellular data system. This is the starting point to the implementation of the process graph theory, which makes a linear approach to overcoming combinatorial explosion possible.

1 INTRODUCTION

The system of cyberworlds is a distributed system. One of the features of cyberworlds is that data dependencies are constantly changing in them. Cyberworlds are more complicated and fluid than any other previous worlds in human history and are constantly evolving. For example, millions of users manage their own blog information every day through Web services on mobile phones, like SNS in Japan, which is considered one of main elements of cyberworlds. At the same time, user requirements for cyberworlds also change and get more complicated as cyberworlds change. If you analyze data using the existing technology in business application development, you have to modify the schema design and application programs whenever schemas or user requirements for output change. That leads to combinatorial explosion, because user requirements, and their combinations and schemas must be specified clearly at the design stage in general business application development. That is a fundamental problem, so we have to reconsider development from the data model level.

Is there a data model that can reflect the changes in schemas and user requirements for output to analyze data in cyberworlds? We believe that the cellular model proposed by one of authors (T. L. Kunii) is the most suitable model. The cellular model based on the Incrementally Modular Abstraction Hierarchy (IMAH) can model the architecture and the changes of cyberworlds and real worlds from a general level to a specific one, preserving invariants while preventing combinatorial explosion (T. L. Kunii and H. S. Kunii, 1999: 19-21). From the viewpoint of IMAH, existing data models are positioned as special cases. For example, UML can model objects at levels below the presentation level, and in the relational data model, a relation is an object at the presentation level which extends a cellular space because it has necessary attributes in which a type is defined, while the processing between relations is based on the set theoretical level. In the object-oriented model, an object is also the object in the presentation level, which extends a cellular space, while the relation between Class is the tree structure, which is a special case of a topological space. An Object in XML is considered a special case of a cellular space which extends a topological space, because an attribute and a value of it are expressed in the same tag format.

In our research, one of the authors (Y. Seki) proposed an algebraic system called Formula Expression as a development tool to realize the cellular model. T. Kodama has actually implemented CDS using Formula Expression (Toshio Kodama, Tosiyasu L. Kunii and Yoichi Seki, 2006: 65-74). In this paper, we have introduced a new concept of a
condition formula and its processing maps into CDS. A condition formula search is a very effective measure when you want to analyze data in cyberworlds without losing consistency in the entire system, since you can search for the data you want without changing application programs, if you employ a condition formula search. In addition, we put emphasis on practical use by taking up an example. First, we explain the cellular model briefly and add a new definition to Formula Expression. (Section 2) Second, we design logical operation as a condition formula generalizing search conditions of users by Formula Expression, and design its processing maps to process a condition formula to each topological space (Section 3) and implement them. (Section 4) We demonstrate the effectiveness of CDS by developing a business application system, thereby abbreviating the process of designing and implementing most application programs. (Section 5) It is a bidding results data search system where the data of files, which schemas differ, are inputted without designing schemas. A more flexible data search is possible by employing a condition formula search in the system.

2 THE CELLULAR MODEL AND FORMULA EXPRESSION

The following list is the Incrementally Modular Abstraction Hierarchy (IMAH) in the cellular model to be used for defining the architecture of cyberworlds and their modeling:
1. the homotopy (including fiber bundles) level
2. the set theoretical level
3. the topological space level
4. the adjunction space level (Fig 1)
5. the cellular space level
6. the presentation (including geometry) level
7. the view/projection level

In this paper, we have added the 3rd bracket [] of (8) in the definition of Formula Expression. The algebraic structure is the following.

[r]×(s+t) = [r]×s+[r]×t, (r+s)×[t] = r×[t]+s×[t]

In this way, if [] is added to a formula and becomes the factor, it behaves like an identifier, since [] is never removed by any map.

3 THE DEFINITION OF LOGICAL OPERATION

3.1 A Condition Formula

If users can specify search conditions, data search will become more functional when searching data from data storage. Here, we introduce the function for specifying conditions defining a condition formula by Formula Expression into CDS. Let propositions P, Q be sets which include characters p, q respectively. The conjunction, disjunction and negation of them in logical operation are defined by Formula Expression as follows:
1) Conjunction
   P∧Q = p×q
2) Disjunction
   P∨Q = p+q
3) Negation
   ¬p = !p

For a detailed explanation of each level, please refer to our earlier paper.
A formula created from these is called a condition formula. Here "!" is a special factor which means negation. Recursivity by () in Formula Expression is supported so that the recursive search condition of a user is expressed by a condition formula. An example is the following.

\[ \neg(P \lor Q) \land ((R \land S) \lor (T \land U)) = !(p+q)(r \times s + t \times u) \]

### 3.2 A Quotient Acquisition Map and a Remainder Acquisition Map

A quotient acquisition map \( f \) is a map that has a term that includes a specified factor, and a remainder acquisition map \( g \) is a map that has a term that doesn’t include a specified factor. These two maps are fundamental in processing a condition formula.

(3.3.) If you assume the entire set of terms to be \( A, B \) and the entire set of factors to be \( C, f: A \times C \rightarrow B \) and \( g: A \times C \rightarrow B \). Arbitrary terms \( r, s, t, u, v, w, x, y \ (A) \) and an arbitrary factor \( p \ (C) \) follow these rules:

\[
\begin{align*}
f: r, p & \varphi \text{ (when } r \text{ doesn’t include } p) \\
f: r \times p \times s, p & \rightarrow r \times s \\
f: r \times (s \times t \times p \times u + v) \times w, p & \rightarrow r \times (s \times t \times p \times u + v) \times w \\
g: r, p & \varphi \text{ (when } r \text{ doesn’t include } p) \\
g: r \times p \times s, p & \rightarrow r \\
g: r \times (s \times t \times p \times u \times v + w) \times p, p & \rightarrow r \times (s \times t \times p \times u \times v + w) \times w \\
g: r \times (s \times t \times p \times u \times v + w), p & \rightarrow r \\
g: r \times [s \times t \times p \times u \times v + w], p & \rightarrow r \times [s \times t \times p \times u \times v + w] \\
\end{align*}
\]

If \( p \) is an identifier, \( f \) (or \( g \)) is usually repeated until \( p \) is not enclosed in a bracket. Simple examples of both maps are shown below.

\[
\begin{align*}
f: a(b(c+d(e+f)))+g)h, d & \rightarrow a(b \times d(e+f))h \\
g: a(b(c+d(e+f)))+g)h, d & \rightarrow a(b \times c+g)h \\
\end{align*}
\]

### 3.3 A Condition Formula Processing Map

A condition formula processing map \( h \) is a map that gets a disjoint union of terms which satisfies a condition formula from a formula. If you assume \( x \) to be a formula and \( x_i \) to be a term which consists of \( x \) (namely \( x_i = x \) and \( p, \neg p, p \times q, p \div q \) to be condition formulas, the images of \( (x, p) \), \( (x, \neg p) \), \( (x, p \times q) \), \( (x, p \div q) \), \( (x, \neg(p+q)) \), and \( (x, \neg(p \times q)) \) by \( h \) are the following:

\[
\begin{align*}
h(x, p) &= i(f(x_i, p)) \\
h(x, p \times q) &= i(f(x_i, p), q) \\
h(x, p \div q) &= i(f(x_i, p), p+q) \\
h(x, \neg p) &= i(g(x_i, p)) \\
h(x, \neg(p+q)) &= i(g(x_i, p), q) \\
h(x, \neg(p \times q)) &= i(g(x_i, p), p+q) + i(g(x_i, p)) \\
\end{align*}
\]

Figure 2: Images from the condition formula processing map.

Fig 2 is images of the map \( h \). Here, \( f \) is a quotient acquisition map and \( g \) is a remainder acquisition map. It is obvious that any complicated condition formula can be processed by the combinations of the above four correspondences. A simple example is shown below.

\[
\begin{align*}
x &= \text{"animal}{\text{color}+\text{size}}(\text{flesheating}{\text{bear}{\text{brown}+\text{big}}}{+}\text{monkey}{\text{brown}+\text{small}}{+}\text{orangutan}{\text{darkbrown}+\text{big}}{+}\text{tiger}{\text{brown}+\text{black}+\text{big}}{+}\text{fox}{\text{brown}+\text{white}+\text{small}}{+}\text{bear}{\text{black}+\text{big}}{+}\text{grasseating}{\text{horse}{\text{white}+\text{brown}+\text{middle}}}{+}\text{koala}{\text{brown}+\text{small}}{+}\text{oat}{\text{white}+\text{small}}{+}\text{hamster}{\text{white}+\text{verysmall}}{+}\text{panda}{\text{black}+\text{white}+\text{big}}{+}\text{zebra}{\text{black}+\text{white}+\text{middle}}{+}\text{giraffe}{\text{yellow}+\text{black}+\text{verybig}}{+}\text{elephant}{\text{gray}+\text{verybig}}{+}\text{mouse}{\text{gray}+\text{verysmall}}")
\end{align*}
\]
Output case 1.

User requirement: "information about a horse and a zebra in x is required"
A condition formula = "horse+zebra"
\[ h(x, \text{horse}+\text{zebra}) \]
\[ = f(x, \text{horse})+g(x, \text{horse}, \text{zebra}) \]
\[ = \text{animal}\{\text{color}+\text{size}\}\text{grasseating}(\text{horse}\{(\text{white}+\text{brown})\}+\text{middle})+\text{zebra}\{\text{black}×\text{white}+\text{middle}\} \]

Output case 2.

User requirement: "information about animals whose size is big or very big and grass-eating is required"
A condition formula = "size (big+verybig) grasseating"
\[ h(x, \text{size(big+verybig)} \text{grasseating}) \]
\[ = f(f(f(x, \text{size}), \text{big+verybig}), \text{grasseating}) \]
\[ = f(f(f(x, \text{size}), \text{big})+g(f(x, \text{size}), \text{big}, \text{verybig}), \text{grasseating}) \]
\[ = \text{animal}\text{size}\text{grasseating}(\text{panda}×\text{big}+\text{giraffe}×\text{very big}+\text{elephant}×\text{verybig}) \]

4 IMPLEMENTATION

This system is a web application developed using JSP and Tomcat 5.0 as a Web server. The client and the server are the same machine. (OS: Windows XP; CPU: Intel Core2 Duo, 3.00GHz; RAM: 3.23Gbyte; HD: 240GB)

Fig. 3 is the flowchart of the algorithm for a quotient acquisition map which is the main function of a condition formula search. Details are abbreviated due to the restriction on the number of pages. In this algorithm, the absolute position of the specified factor by the function of the language and the term including the factor are acquired first. Next, the nearest brackets of the term are acquired and because the term becomes a factor, a recursive operation is done.

5 CASE STUDY: A BIDDING RESULTS DATA SEARCH SYSTEM

5.1 Outline

We have developed a business application system using CDS for searching bidding results data for public construction projects. Many of the data files were downloaded in CSV format from the official website of each bureau in the Ministry of Land, Infrastructure, Transport and Tourism (MLIT, http://www.mlit.go.jp/chotatsu/kekka/kekka.html) in Japan. The feature of the files is that their schema changes little from month to month or from bureau to bureau. Once you convert the CSV data files to formulas in CDS, you can unify them into a data storage file (.txt) by the function of a disjoint union +. After that, a user can search for the data she/he wants from the data storage by creating a condition formula. This system is actually being used in Maeda Corp. which one of authors (T. Kodama) belongs to.

5.2 The Space Design

We design a formula for the spaces as follows.

\[ \Sigma_{\text{file}};\Sigma_{\text{code}}\{[\Sigma_{\text{attribute}};\Sigma_{\text{value}}]\} \]

\text{file}: a factor which expresses a file name
\text{attribute}: a term which expresses an attribute name of \text{file}
5.3 Data Conversion and Data Input

In this subsection, we simplify the input data without losing generality. Let the CSV data of Figure 5 be bidding results from May, 2007 in the Tohoku bureau. First, convert the downloaded CSV data to a formula (formula 5.3-1) as a cell space and add it to the data storage file.

Formula 5.3-1.


In this way, you can add data to the data storage after converting it to a formula for a cell space using + function.

Next, add the CSV data from June in the Kanto bureau, which schema is slightly different from that in formula 5.3-1, convert it to a formula (formula 5.3-2) in the same way and add it to the data storage file using + function.

Formula 5.3-2.


In this way, you can add data to the data storage after converting it to a formula for a cell space using + function.
+ function. In doing this, you don’t need to consider differences in schema at all. In the same way, you can add data from another organization, which schema is completely different from others’, as a formula for a cell space to the data storage file.

### 5.4 Data Conversion and Data Input

When you search for data you want, you create condition formulas according to requirements and get an image of the formula in data storage by the condition formula processing map \( h \), you can get the data you want. Examples and figures (Fig 7,8) are shown below.

If you want to search for data for "construction projects of Company C1 or C2 and for WTO (World Trade Organization) property", you make the condition formula "(C1+C2)WTO", and get the image of formula 5.3-2 by the condition formula processing map \( h \).

\[
\begin{align*}
\text{h (formula 5.3-2, (C1+C2)WTO)} &= \text{MayOf2007InTohoku\_code[bureau\_name+project\_name+bid\_date+contract\_date+project\_kind+bidding\_company+bidding\_price]}[7\{\text{tohoku+B+12/06/2007+13/06/2007+general\_construction+general\_bid\_WTO+C1+50000000}\};8\{\text{tohoku+B+12/06/2007+13/06/2007+general\_construction+general\_bid\_WTO+C2+51000000}\}];=\text{MayOf2007InKanto\_code[bureau\_name+project\_name+place+bid\_date+contract\_date+delivery\_month+project\_kind+bidding\_company+bidding\_price]}[1\{\text{kanto+C+p1+15/06/2007+17/07/2007+5+general\_construction+general\_bid\_WTO+C1+800000000}\};2\{\text{kanto+C+p1+15/06/2007+17/07/2007+5+general\_construction+general\_bid\_WTO+C2+780000000}\}];
\end{align*}
\]

Next, if you want to search for data for "construction projects in the Kanto bureau which are not for WTO", you create the condition formula "kanto\_!WTO", and get the image of formula 5.3-2 by the condition formula processing map \( h \).

\[
\begin{align*}
h (\text{formula 5.3-2', (C1+C2)}WTO) &= \text{MayOf2007InKanto\_code[bid\_date+contract\_date+project\_kind+bidding\_company+bidding\_price]}[1\{15/06/2007\};2\{15/06/2007\};3\{15/06/2007\};4\{15/06/2007\};5\{15/06/2007\};6\{15/06/2007\};7\{15/06/2007\};8\{16/06/2007\};9\{16/06/2007\};10\{16/06/2007\};11\{16/06/2007\};12\{16/06/2007\};13\{16/06/2007\}];
\end{align*}
\]

5.5 Considerations

When a business application system like the one above is developed in the existing way, user requirements are analyzed first. Next, the system, schemas and application programs are designed according to requirement analysis. Then, implementation and testing are done. The fundamental development process is changed if CDS is used.

1. Schema Design and Data Input

It is almost impossible for a database designer to design schema of this application system since she/he cannot predict the changes in schema of MLIT bidding results data. And whenever a new file which schema is different from that already designed appears, it is actually impossible to modify the schema design and application programs or to develop data conversion programs. If you employ
CDS in the development of this application system, you don’t have to worry about the above problems. This is because the concept of the disjoint union + of the cellular model is supported in CDS, so that you can add the data which schema are different to the data storage one after another, if you only have to convert the data to formulas of CDS.

2. Data Output
Data output design and application programs for data output have to be done during application system development, and they have to be modified when there is a new user requirement for output which was not expected in the user requirement analysis. This can be costly. But if you use CDS in the development, a user only has to create a condition formula according to a user requirement for output. This is because user requirements can be generalized by condition formulas of CDS.

3. Processing Speed
Detailed benchmark tests have not been conducted yet, but when we actually tried this system, the output processing speeds of 500 records and 1,000 records from more than 200,000 records were 3.2s and 6.7s respectively. This system is considered practical for analyzing business data on a client PC.

6 RELATED WORKS

The distinctive features of our research are the application of the concept of topological process, which deals with a subset as an element, and that the cellular space extends the topological space, as seen in Section 2. Relational OWL as a method of data and schema representation is useful when representing the schema and data of a database (Takashi Washio and Hiroshi Motoda, 2003: 59-68), but it is limited to representation of an object that has attributes. Our method can represent both objects: one that has attributes as a cellular space and one that doesn’t have them as a set or a topological space. Many works applying other models to XML schema have been done. The motives of most of them are similar to ours. The approach in (Giovanna Guerrini, Marco Mesiti, Daniele Rossi, 2005: 39-44) aims at minimizing document revalidation in an XML schema evolution, based on a part of the graph theory. The X-Entity model (Bernadette Farias Lósio, Ana Carolina Salgado and Luciano do Rêgo Galvão, 2005: 39-44) is an extension of the Entity Relationship (ER) model and converts XML schema to a schema of the ER model. In the approach of (N. Routledge, L. Bird and A. Goodchild, 2002: 157-166), the conceptual and logical levels are represented using a standard UML class and the XML represents the physical level. XUML (HongXing Liu, YanSheng Lu, Qing Yang, 2006: 973-976) is a conceptual model for XML schema, based on the UML2 standard. This application research concerning XML schema is needed because there are differences in the expression capability of the data model between XML and other models. On the other hand, objects and their relations in XML schema and the above models can be expressed consistently by CDS, which is based on the cellular model. That is because the tree structure, on which the XML model is based, and the graph structure, on which the UML and ER models are based, is special cases of a topological structure mathematically. Entity in the models can be expressed as the formula for a cellular space in CDS. Moreover, the relation between subsets, as we showed in 3.2, cannot in general be expressed by XML. Although CDS and the existing deductive database look alike apparently, the two are completely different. The deductive database (Q. Kong, G. Chen, 1995: 973-976) raises the expression capability of the relational database (RDB) by defining some rules. On the other hand, CDS is a proposal for a new tool for data management and has nothing to do with the RDB.

7 CONCLUSIONS AND FUTURE WORKS

In this paper, we have developed a condition formula search as an important function of CDS. Using this function of CDS, you can search for data you want from formulas as data storage by creating a condition formula according to user requirements, so that you don’t have to analyze user requirement for output in typical business application development. The point we should emphasize for future work is that the search condition of a user as well as data for input/output is expressed as a formula. This certainly brings the system which is developing, including user requirements recursively. This will be connected to the implementation of a process graph.
It is the next step where a situation as a node is transferred to the next situation selecting a path as an edge. Implementation has been difficult up to the present time because there is no tool to realize it, although one of authors (T.L. Kunii) outlined the plan many years ago. The appearance of Formula Expression will enable it in the near future. If we implement the process graph by developing CDS as future work, automation of business application development will be done. We believe that CDS brings great social impact, changing existing development fundamentally. Our research is still in its infancy, but it is progressing every day. We are collaborating with companies and universities worldwide.

REFERENCES


ANALYZING DECENTRALIZED GOVERNABILITY OF BUSINESS PROCESSES BY EXTENDED PETRI NETS AND MODAL LOGICS

Takashi Hattori
NTT Communication Science Laboratories, 2-4, Hikaridai, Seika-cho, Keihanna Science City, Kyoto, Japan
takashi_hattori@cslab.kecl.ntt.co.jp

Hiroshi Kawakami, Osamu Katai, Takayuki Shiose
Graduate School of Informatics, Kyoto University, Yoshida Honmachi, Sakyo, Kyoto, Japan
{katai, kawakami, shiose}@i.kyoto-u.ac.jp

Keywords: Decentralized Control, Discrete Event System, Petri Net, Modal Logic.

Abstract: We introduce a novel notion of decentralized governance structure of event-driven processes together with the notions of their behavioral and structural correctness. The ways for attaining correct process behavior, as well as the notion of decentralized governability, are examined based on temporal logical analyses of process behavior via Petri net representations of process structures. Also, the deontic and temporal logical prescriptions of normative constraints (tasks) on the processes are introduced that are then translated into extended hierarchical Petri net structures. The conflicts among these tasks are examined on this hierarchical structure.

1 INTRODUCTION

As e-businesses grow, users have gained power in obtaining information and combining various service applications. For instance, if a user decides to go on a trip, s/he may use an access map, train information, a hotel search, and a defrayment service. Thus, users are nowadays active, unlike those of the past that only passively followed the line prepared by service providers. In other words, a single service provider is not as powerful now as before, and processes are not as centralized as before.

This paper proposes a theoretical framework for such decentralized multi-agent (in this case, users and service providers) systems that can represent agents’ behavior and their policies or control rules. The behaviors of agents are represented by a Petri net (Peterson, 1981), which offers rich mathematical analysis, and introducing modal logics (Hughes and Cresswell, 1968) enables us to represent policies and control rules. It is known that a Petri net is conventional and now its successors, e.g., a Unified Modeling Language (UML) (Saldhana and Shatz, 2000), are popular tools for system modeling. Applying a Petri net is still a hot topic for those researchers that place emphasis on checking the behavior of system design (Hu and Shatz, 2004). Modal logics, e.g., temporal and deontic logics, are also known as conventional theories. They have been applied to several systems (Blackburn et al., 2006) and their theoretical studies still progress (Nute, 2004).

For further development of e-business, decentralized governance is inevitable. Growing numbers of services and applications may lead to overlap of services, which may sometimes cause interference. Our framework enables us to check the existence of interference among services’ control and user’s policies. Furthermore, it shows us how to eliminate the interference.

The rest of this paper consists of the following sections. Section 2 introduces the outline of the framework. In this framework, components of a target system are classified into two portions from the viewpoint of whether it is a prefixed structural element or an element that may change over time like users’ policies and service providers’ control rules. The former portion is encoded into a conventional Petri net. The latter half is first represented by modal logic formulae, which are then translated into extended Petri nets that we call “task unit graphs.” The two portions are then integrated into a single extended Petri net. In section 3, the framework introduced in section 2 is applied to the modeling of decentralized governance problems. Based on the model, we discuss the correctness of process behavior and decentralized/decentralized governability in sections 4 and 5.
2 MODELING SYSTEMS BY PETRI NET AND MODAL LOGIC

This section proposes a method for modeling decentralized systems based on a kind of Petri net and modal logic. Hereafter, we present the procedure of system modeling by using an example of a travelers’ decision on itineraries with the assistance of e-applications.

Example System. Assume that in the near future, many e-applications will work as sophisticated agents, and help users by cooperating with each other. Now a traveler has arrived at Porto station. S/he enters an internet cafe, and launches the following agents: e-landmark map agent (A1), e-hotel search (A2), e-train connection information (A3), and an e-defrayment system (A4).

S/he has two itineraries to decide on. One is finding a hotel to stay at and a way to get there by train. The other is finding sightseeing spots and how to access them. In the former mission, s/he first searches for a hotel by using A2, which submits the hotel’s location to A3, as well as her/his current location. A3 determines the route between Porto and the nearest station to the hotel. After the route is decided, defrayment is executed by A4. In the latter mission, s/he decides on two scenic sites s/he wants to visit with the help of A1, which hands over the location of the sites to A3 and defrayment is again executed by A4. Anyway, A3 requires two locations and searches for the optimal route between them.

2.1 Petri Net Representation of System’s Event-Driven Aspect

A Petri net is known as a conventional representation scheme for modeling a physical structure and event-driven behavior of discrete event systems (Karatkevich, 2007). It is also known that a k-bounded standard Petri net can be translated into an equivalent 1-bounded Petri net. We employ a 1-bounded one called the condition/event system (C/E system) (Reisig, 1982) where a transition can only fire if all “its output places” are empty.

For instance, the example described above is the case where the upper-bound of “the number of stored locations” is two (2-bounded). Thus it can be modeled as a C/E system as shown in Fig. 1, where a token in place Pj means the following:

- P1: Both A1 and A2 are idling,
- P2: A1 is working,
- P3: A2 is working,
- P4: cache memory of A3 stores a location,
- P5: second memory of A3 stores a location,
- P6: A4 is idling,
- P7: A4 is working.

Each of the transitions τ2 and τ4 means a submission of the location of a hotel or a landmark, τ3 means the “data transfer from the cache to the second memory” and “flushing cache,” and the firing of τ6 makes A3 search for a train connection between two locations and give train fees to A4.

![Figure 1: Petri net representation of a decentralized discrete event system.](image)

When we correspond a place of a C/E system to a proposition, we can represent the true/false value of the proposition by putting/removing a token in/from the place. In this case, each transition leads the value alteration of the proposition. For instance, in Fig. 1, the firing τ6 leads P4, P5, P6 to turning from true to false, and P7 to turning from false to true.

2.2 Modal Logic Representation of Tasks

Next, we represent the tasks that state “when the focused state should be true” as propositions by introducing temporal and deontic logic.

2.2.1 Temporal Modalities

A temporal logic is given by the propositional logic, modal operators, and an axiom system. This paper employs the following modal operators:
TA: A will be true at the next state S₁.
GA: A will be true from now on S₀, S₁, S₂, ….
FA: A is true at So, or will be true at some time in the future Sᵢ(i > 0),
A t B: B is true at S₀ or A will be true from now on until the first moment when B will be the case,

where A, B denote logic formulae, and S₀/Sᵢ (i > 0) mean current/future states (worlds) respectively.

Axiom systems of temporal logic vary depending on the viewpoint of time. This paper employs one of the discrete and linear axiom systems K₅₅ (Katai and Iwai, 1983), which is an extension of the minimal axiom system K₅ (Rescher and Urquhart, 1971). Introducing ⊢ (yesterday) as the mirror image of ⊢ (tomorrow), the axiom system claims that ⊢ γ a ≡ ¬γ a, ⊢ γ a ≡ ¬γ a, and ⊢ γ ⊢ a ≡ γ ⊢ a. Introducing ⌜ (since) as the mirror image of ⊢ (until), ⊢ γ a ≡ a ⊢ ⊥, where ⊥ denotes the contradiction, and ⊢ γ a ≡ ¬γ a. ⊢ a is rewritten as

⊢ a ⊢ B ≡ B ∨ (A ∧ ⊢ (a ⊢ B)),
⊢ a ⊢ B ≡ B ∨ (A ∧ ⊢ (a ⊢ B)),

where A, B and C denote logic formulae.

Regarding state transitions in the future, there are two aspects, i.e., b (branching) and l (linear), thus the modes γ and ⊢ are more precisely defined as (Katai, 1981):

γ a: A will necessarily be persistent,
γ a: A will possibly be persistent,
FA: A will possibly be the case,
FA: A will necessarily be the case.

Figure 2 illustrates those modes where each circle denotes a state, each arc denotes a state transition, and the black circles mean that the state holds A.

Furthermore, ⊢ a also can be branching (uᵢ) or linear (uᵢ) as shown in Fig. 3. In the figure, each black circle means that B is true in that state, and the letter A means that A is true in that state. Among them, the following relations are established: γ a B ≡ A ⊢ (B ∧ ¬B) ≡ ¬γ a B, γ a B ≡ A ⊢ (B ∧ ¬B) ≡ ¬γ a B.

2.2 Deontic Modalities

Understanding the system’s behavior by temporal logic is of an “objective” view of the focused proposition. To represent our “subjective” intention or purpose, such as how the propositions should behave, i.e., the control rule or task, we introduce deontic modalities:

□ A: A is obligatory, □ A: A is permitted.

The axiom system we adopt here for □ and □ is that of SDL (standard deontic logic), which defines □ A ≡ ¬□ ¬A, and claims ⊢ □ A ⊢ □ A, and ⊢ □ □ (A ⊢ B) ⊢ (□ A ⊢ □ B).

Some control rules and specifications of systems can be translated into the combinations of temporal and deontic modes by using “translation templates” such as

□ A: A has to be true in the future,
□ A: A can be always true.

They correspond to alethic modes □ A and □ A respectively.

2.3 Network Representation of Tasks

We translate the task represented by modal logic into an extended Petri net, which we call a “task unit graph,” by introducing four types of special arcs shown in Fig. 4.

Figure 4: Special arcs for control of transition firing.
These arcs are placed from a place to a transition. They function whenever the place holds a token and the transition satisfies the firing condition, but they differ from regular arcs of the conventional Petri net on the following points. First, they do not transfer tokens from places to arcs. Next, if there are multiple special arcs from the same place, all of them are activated simultaneously. As a result, simultaneous firing of multiple transitions is permitted at the same state.

Figure 5 shows examples of task unit graphs, and these net representations are derived by a systematic analysis of logical representations of tasks. For instance, consider a task \( O(A \cup B) \). A has to be true from now on until \( B \) will be the case. If \( B \) is the case at \( S_0 \), this task is accomplished, else if \( \neg A \And \neg B \) at \( S_0 \), this task cannot be accepted due to Eq. (1). If \( A \And \neg B \) at \( S_0 \), \( A \) should be maintained and \( O(A \cup B) \) also has to be the case at \( S_1 \). In each case, \( O(A \cup B) \) at \( S_0 \) prohibits the alteration from \( A \to \neg A \), so an arc of “prohibition of firing” is placed from the place of \( O(A \cup B) \) to the “transition of the alteration from \( A \to \neg A \)”.

3 DECENTRALIZED SYSTEM STRUCTURE

This paper defines the characteristics of decentralized systems as

- each agent has its own territory,
- each control task is given to one of these agents,
- a task can control a transition within the territory of the agent to whom the task is given.

Figure 1 is an example of decentralized system. It consists of four agents who are in charge of managing the sub tasks denoted by thick broken circles in Fig. 1.

Assume that the traveler has her/his own policy for using e-agents and adopts them as control rules in order to fulfill her/his needs such as:

- PP1: defrayment must not be done simultaneously with submission of a landmark/hotel location;
- PP2: hotel search is always followed by a landmark search;
- PP3: submission of either a hotel or a landmark is accepted only if the first cache of the train search is empty;
- PP4: the system should request a hotel submission some time in the future before the end of a set of executions.

Each task associated with each agent is activated by the firing of the corresponding transition in its territory, e.g., in this case we have the following tasks represented in temporal deontic logical forms,

\[
\begin{align*}
PP0: & \quad \tau_0 \text{ activates } O_2(P_1 \cup P_2) \\
PP1: & \quad \tau_0 \text{ or } \tau_1 \text{ activates } O_2(P_3 \cup P_4) \\
PP2: & \quad \tau_0 \text{ activates } O_2(\neg P_4 \cup P_5) \\
PP3: & \quad \tau_0 \text{ or } \tau_4 \text{ activates } O_1(P_1 \cup (\neg P_2)) \\
PP4: & \quad \tau_1 \text{ activates } O_2 F P_3 \\
\end{align*}
\]

where \( O_i \) stands for the obligation for agent \( A_i \).

Not every task corresponds to a specific transition. Some tasks are translated into logical forms that are not activated by a transition but are always activated. For example, a rule

\[
PP0: \quad \text{Once the defrayment process is finished, agent } \ A_2 \text{ should not submit a hotel location until the cache memory flushes its contents to the memory, is resident and translated into } \\
O_2 G (\{ (\neg P_4) \cup (\neg P_4 \And P_5) \} \cup (\neg P_5)).
\]

Defining \( Q \equiv \neg P_4 \And P_5 \). \( H \equiv \neg P_4 \And Q \). PP0 can be represented as \( O_2 G (H S(\neg P_4)) \), which derives an extended Petri net representation as shown in the left part of Fig. 6 (Katai, 1981).

It consists of module nets reflecting its subtasks, and they are joined with linkage relations prescribing concurrent (simulations) firing of linked relations shown in Fig. 4 (d). It should be noted that the transitive closure of these linkage relations links the task in analysis with the target system, which is shown in the middle part of Fig. 6. In the figure, task unit graphs showing \( \sigma \) and \( \And \) (conjunction) are employed. Their general types are defined as shown in Fig. 7.

4 CORRECTNESS OF SYSTEM AND ITS BEHAVIOR

4.1 Behavioral Corrections

A system can be regarded to behave correctly iff it satisfies the following conditions:
Definition 1 (Correctness of State Transition Sequence). A sequence of state transition is correct iff the following cases do not occur along with this infinite sequence:

1. Any proposition $A$ does not occur just after the state where $O_i \triangleright A \ (O_i \triangleright A, O_i (A U B))$ is given to an agent $i$.
2. $A$ never happens to the case after the state (until $B$ is the case) where $O_i \triangleright F A \supset F b A$ is given.

4.2 System Correctness and its Characterization

We have two kinds of system corrections as follows:

Definition 2 (Strong Correctness of System). A system is strongly correct iff any state transition sequence generated by the system (system behavior) is correct. Namely, there is no need to control the system.

In terms of temporal or deontic logical expressions, the above is characterized as:

\[
\begin{align*}
&\diamond_b (\neg C), \\
&\diamond_b (O_i \triangleright A \supset \triangleright_b A), \\
&\diamond_b \{O_i (\neg (A U B)) \supset (\neg (A U b B))\},
\end{align*}
\]

where $C$ is a state of contradiction.

We have practically more important and weaker notions of system correctness as follows:

Definition 3 (Weak Correctness - Centralized Governability). A system is called weakly correct iff we can extend an arbitrary generated state transition sequence so that it is correct by appropriately executing the firing of permitted (legal) transitions. This can be characterized as

\[
\begin{align*}
&\diamond_b (\neg C), \\
&\diamond_b (O_i \triangleright A \supset \triangleright_b A), \\
&\diamond_b \{O_i (\neg (A U B)) \supset (\neg (A U b B))\}.
\end{align*}
\]

In terms of the state transition diagram, the current notion of strong correctness can be characterized as:

Theorem 1 (Strong Correctness of System). A system is strongly correct iff the following hold:

(a) for an arbitrary terminal state of its state transition diagram, there is no task associated with it of the form of $O_i \triangleright A$, and if $O_i \triangleright A$ is present there, then $A$ is also present on that state, and if $O_i (\neg (A U B))$ is present there, then both $\neg A$ and $\neg B$ are also there;

(b) for an arbitrary cycle (circuit) of its state transition diagram, the following hold:
(b.1) if $\varphi A$ is present at a state in the cycle, then $A$ is also present at possibly another state in the cycle;

(b.2) if $\varphi (\neg(A \cup B))$ is present on a state $s$ in the cycle, then there is also state $s'$ such that $\neg A$ holds on $s'$ and there is no state between $s$ and $s'$ at which $B$ holds.

For characterizing weak correctness, we introduce the notion of condensation of directed graphs by “strong components” that are defined as their bidirectionally connected maximal subgraphs (Harary et al., 1965).

**Theorem 2 (Weak Correctness of System).** A system is weakly correct iff the following hold:

(c) the same as condition (a) in Theorem 1;

(d) for every terminal strong component of its state transition diagram the following hold:

(d.1) if $\varphi A$ is present in a state at the component, there is a state (and possibly another) on which $A$ holds;

(d.2) if $\varphi (\neg(A \cup B))$ is present in a state $s$, then there is a state $s'$ on which $\neg A$ holds and there is a path joining $s$ and $s'$ along which $B$ never holds.

## 5 DECENTRALIZED GOVERNABILITY OF PROCESSES

### 5.1 Method of System Correction

From the above results, we will have two ways of making an arbitrary system to behave correctly:

(i) to make the system strongly correct,

(ii) first to make the system weakly correct and then to control it so that its behavior (generated state transition sequence) becomes correct.

In the first approach, there is no more need to control it, i.e., any state transition sequence yielded from it is surely correct. In the latter approach, weak correctness itself is merely a precondition on governability and there is still need for supplementary control on permitted transitions. In other words, weak correctness guarantees the possibility of this supplementary control. In this paper we will pursue the latter approach, which seems to be of more practical importance than the former.

### 5.2 System Correction in Terms of State Transition Diagram

It can be readily seen that the following modifications on state transition diagrams are necessary for making systems be weakly correct:

(ii.1) remove the terminal states from the diagram at which either a task

$\varphi A$ is present, or

$\varphi A$ and $\neg A$ are present, or

$\varphi (\neg(A \cup B))$ is present and at least one of $\neg A$ or $\neg B$ is absent;

(ii.2) remove the terminal strong components, which include

$\varphi A$ but $\neg A$, or

$\varphi (\neg(A \cup B))$ is present but $\neg A$, or

$\varphi (\neg(A \cup B))$ at a state $s$ and $\neg A$ at a state $s'$ such that any path from $s$ to $s'$ includes a state at which $B$ does not hold.

The above operations on state transition diagrams need to be applied repeatedly because removal of terminal states or strong components will yield different terminal states and strong components. The operations proceed until there is no need for them. If we still have remaining states in the diagram, the system is modified to be weakly correct.

### An Example of Conflict Detection

The typical conflicts are observed among tasks. Figure 8 shows the diagram of a portion of the sequence of state transitions of the target system with the initial state S0, which holds $P_1$ and $P_6$, and is in charge of tasks PP0, 1, 2, 3, and 4. Table 1 shows the markings of each state where “•” means a normal token and “○” means an active token, which constraints other tokens.

![State transition diagram of the system in table 1](image)

The terminal state S5 is removed from the diagram by the operation (ii.1) since $\varphi P_5$ is required but $\neg P_5$ at the state. As a result, there is no “terminal strong component” in the diagram and the system becomes weakly correct.

The conflict in the state S5 can be detected by tracing synchronized firing linkages (broken lines in Fig. 9) as mutual interferences among tasks. In state
S5, place $O_1 P_1 \cup (\neg P_4)$ has a token, which prohibits firing of $\tau_1$ and $\tau_3$. On the other hand, the token in place $O_2 F P_3$ requests firing of $\tau_3$. Therefore, there is a conflict of firing $\tau_3$ in state S5. The only way to resolve this conflict is turning $P_4$ to $\neg P_4$, which leads the token in $P_1 \cup (\neg P_4)$ to free. But the establishment of $\neg H$ in state S5 prohibits turning $P_4$ to $\neg P_4$ by tracing synchronized firing linkages from $O \mathcal{G} (H S (\neg P_7))$. As a result, this conflict cannot be resolved unless $\neg H$ turns to $H$.

5.3 Derivation of Control Rules

The above modifications on transition diagrams can be translated into control actions on the extended Petri net systems. The removal of terminal states becomes (ii.1’) prohibit the firing of transitions just before (leading to) the removed components.

(ii.2’) prohibit the firing of transitions just before (leading to) the removed components.

It should be noted that we need supplementary control actions over weakly correct systems for making their behavior correct. More precisely, by referring to conditions (d.1) and (d.2) in Theorem 2, we need the following operations:

(d.1’) if we arrive at a state where a task of the form $O \mathcal{G} A$ is present, then we must eventually (surely in the future) arrive at a state where $A$ is realized.

(d.2’) if we come to a state where $O_1 (\neg (A \cup B))$ is present, then we must eventually arrive at a state where $\neg A$ holds by going through states at which $B$ is not the case.

5.4 Decentralized Governability

In the above control operations, we have to consider the decentralized nature of systems, i.e., each agent
having its own territory over which it has control. Namely, we should set the following:

**Condition (System Decentralization).** Each prohibition of firing of a transition (ii.1’) or (ii.2’) must be caused by a task that is given to an agent whose territory includes this transition.

Thus, all the prohibition operations along with the course of deriving a weakly correct system should be subject to this condition. It should be noted that there may be various ways of deriving weakly correct systems, and only a portion of them may satisfy the above condition. Hence, it is not easy to verify the following property of an arbitrarily given decentralized system.

**Definition 4 (Decentralized Governability).** A system is called “decentralized controllable” if there exists a sequence of operations (ii.1’) and (ii.2’) in which all the prohibitions of transition firing are in accordance with the above condition on system decentralization.

6 CONCLUSIONS

We have introduced the notions of decentralized governance of event-driven processes, their behavioral and structural correctness, and centralized and decentralized control for attaining correct behavior, as well as that of decentralized governability. Also the method of deriving correct behavioral processes is shown. The basic framework we adopted was temporal and deontic logical analyses of process behavior and Petri net representations of process structures.

The hierarchical decomposition of tasks elucidates the governance (control) structure of tasks over the event-driven process described by Petri net systems. The control flows descend the hierarchy, while the flows of information reporting the changes of object systems ascend the hierarchy. These flows go along the fire synchronization arcs. Also the conflicts between the control flows are elucidated. These analyses and methods are expected as a basis for treating complex business processes that are subject to high reliability and credibility under complicated decentralized governance structures.

REFERENCES


AN ARCHITECTURE FOR DYNAMIC INVARIANT GENERATION IN WS-BPEL WEB SERVICE COMPOSITIONS

M. Palomo Duarte, A. García Domínguez and I. Medina Bulo
Department of Computer Languages and Systems, E.S.I., University of Cádiz
C/ Chile s/n. 11002 Cádiz, Spain
manuel.palomo@uca.es, antonio.garciadominguez@alum.uca.es, inmaculada.medina@uca.es

Keywords: Web Services, service composition, WS-BPEL, white-box testing, dynamic invariant generation.

Abstract: Web services related technologies (especially web services compositions) play now a key role in e-Business and its future. Languages to compose web services, such as the OASIS WS-BPEL 2.0 standard, open a vast new field for programming in the large. But they also present a challenge for traditional white-box testing, due to the inclusion of specific instructions for concurrency, fault compensation or dynamic service discovery and invocation. Automatic invariant generation has proved to be a successful white-box testing-based technique to test and improve the quality of traditional imperative programs. This paper proposes a new architecture to create a framework that dynamically generates likely invariants from the execution of web services compositions in WS-BPEL to support white-box testing.

1 INTRODUCTION

Web Services (WS) and Service Oriented Architectures (SOA) are, according to many authors, one of the keys to understand e-Business in the early future (Heffner and Fulton, 2007). But as isolated services by themselves are not usually what customers need, languages to program in the large composing WS into more complex ones, like the OASIS standard WS-BPEL 2.0 (OASIS, 2007), are becoming more and more important for e-Business providers (Curbera et al., 2003).

There are two approaches to program testing (Bertolino and Marchetti, 2005): black-box testing is only concerned about program inputs and outputs, while white-box testing takes into account the internal logic of the program. The latter produces more refined results, but it requires access to the source code. However, WS-BPEL presents a challenge for traditional white-box testing, due to the inclusion of WS-specific instructions to handle concurrency, fault compensation or dynamic service discovery and invocation (Bucchiarone et al., 2007).

Automatic invariant generation (Ernst et al., 2001) has proved to be a successful technique to assist in white-box testing of programs written in traditional imperative languages. Let us note that, throughout this work, to be consistent with related work (Ernst et al., 2001) we use invariant and likely invariant in its broadest sense: properties which hold always or in a specified test suite at a certain program point, respectively. We consider that dynamically generating such invariants, backed by a good test suite, can become an interesting help in WS-BPEL white-box testing.

We propose a new architecture for a framework to dynamically generate likely invariants of a WS-BPEL composition from actual execution logs of a test suite running on a WS-BPEL engine. These invariants can be an interesting aid for white-box testing the composition.

The rest of the paper is organized as follows. First, we explain the particularities of WS compositions built with WS-BPEL. The next section shows how dynamic generation of invariants can successfully help in WS composition white-box testing. In the following main section we introduce our proposed architecture, discuss how to solve some technical problems that might arise during its implementation, and show an example of the expected results. Finally, we compare our proposal with other alternatives and present some conclusions and an outline of our future work.
2 WS COMPOSITIONS AND WS-BPEL

The need to compose several WS to offer higher level and more complex services to suit customers requirements was detected and satisfied by the leading companies in the IT industry with the non-standard specification BPEL4WS, which was submitted to OASIS in 2003 (OASIS, 2003). Soon after, OASIS created the WS Business Process Execution Language Technical Committee to work in its standardization, releasing the first standard specification, WS-BPEL 2.0, in 2007 (OASIS, 2007).

Standardization has been an important milestone for WS-BPEL wide adoption by SOA leading tools, being a key interoperability feature offered by many e-Business systems (Domnguez Jimnez et al., 2007). This way, companies providing services can take advantage of the new possibilities that this programming-in-the-large technology allows for: concurrency, fault recovery and compensation or dynamic composition using loosely coupled services from different providers selected through several criteria (such as cost, reliability or response time).

WS-BPEL is an XML-based language using XML Schema as its type system. WS-BPEL specifies service composition logic through XML tags defining activities like assignments, loops, message passing or synchronization. It is independent of the implementation and platform of both the service composition system and the different services used in it. There are other W3C-standardized XML-related technologies (W3C, 2008) at its foundation:

- XPath. allows us to query XML documents in a flexible and concise way. Although the latest version of the language is XPath 2.0, WS-BPEL uses by default XPath 1.0.
- SOAP. is an information exchange protocol commonly used in WS. It is platform independent, flexible and easy to extend.
- WSDL. is a language for WS interface description, detailing the structure every message to be exchanged, the interfaces and locations of the services offered, their bindings to specific protocols, etc.

There are several technologies that can extend WS-BPEL. They are usually referred to as the WS-Stack (Papazoglou, 2007). One of the most interesting is UDDI, a protocol that allows maintaining repositories of WSDL specifications, easing the dynamic discovery and invocation of WS and the access to their specifications. UDDI 3.0 is an OASIS Standard (OASIS, 2008).

But the inherent dynamic nature of these technologies also implies new challenges for program testing, as most traditional white-box testing methodologies cannot be directly applied to this language.

3 WS-BPEL WHITE-BOX TESTING

WS composition testing is one of the challenges for its full adoption in industry in the forthcoming years. The dynamic nature of WS poses a challenge for testing, having to cope with aspects like run-time discovery and invocation of new services, concurrency and fault compensation.

Little research has been done on applying white-box testing directly on WS-BPEL code. Main approaches (Bucchiarone et al., 2007) create simulation models in testing-oriented environments. But simulating a WS-BPEL engine is very complex, as there is a wide array of non-trivial features to be implemented, such as fault compensation, concurrency or event handlers. In addition, it is sometimes necessary to translate the code of the WS-BPEL composition to a second language to check its internal logic.

In case any of these features is not properly implemented, compositions would not be accurately tested. So we consider that this is an error-prone process, as it is not based on the actual execution of the WS-BPEL code in a real environment, that is a WS-BPEL engine invoking actual services.

Therefore, we propose using dynamically generated invariants from actual execution traces as a more suitable approach.

3.1 Using Invariants

An invariant is a property that holds at a certain point in a program. Classical examples are function pre-conditions and post-conditions, that is, assertions which hold at the beginning and end of a sequence of statements. There are also loop invariants, which are properties that hold before every iteration and after the last one.

Let us illustrate these concepts with a simple example. Suppose that we want to sum all the integers from 0 to a given natural $n$. In this case, we could define this simple algorithm:

1. $r ← 0$
2. From $i ← 1$ to $n$: $r ← i + r$
3. Return $r$
This algorithm has the pre-condition \( n > 0 \) in step 1, as \( n \) is natural by definition. Looking at the loop in step 2, we can tell that the loop invariant \( r = \sum_{i=0}^{j-1} i \) holds before every iteration. Since in step 3 we have exited the loop, we will have \( i = n + 1 \), which we can substitute in the previous loop invariant to formulate the post-condition for step 3 and the whole algorithm, \( r = \sum_{i=0}^{n} i \). From this post-condition we can tell that the algorithm really does what it is intended to do.

Manually generated invariants have been successfully used as above to prove the correctness of many popular algorithms to this day. Nonetheless, their generation can be automated. Automatic invariant generation has proved to be a successful technique to test and improve programs written in traditional structured and object-oriented programming languages (Ernst et al., 2001).

Invariants generated from a program have many applications:

**Debugging.** An unexpected invariant can highlight a bug in the code which otherwise might have been missed altogether. This includes, for instance, function calls with invalid or unexpected parameter values.

**Program Upgrade Support.** Invariants can help developers while upgrading a program. After checking which invariants should hold in the next version of the program and which should not, they could compare the invariants of the new version with those of the original one. Any unexpected difference would indicate that a new bug had been introduced.

**Documentation.** Important invariants can be added to the documentation of the program, so developers will be able to read them while working on it.

**Verification.** We can compare the specification of the program with the actual invariants obtained to see if they satisfy.

**Test Suite Improvement.** A wrong likely invariant dynamically generated, as we’ll see in next section, can demonstrate a deficiency in the test suite used to infer it.

### 3.2 Automatic Invariant Generation

Basically, we there are two approaches when generating invariants automatically: static and dynamic.

Static invariant generators (Bjørner et al., 1997) are most common: invariants are deduced statically, that is, without running the program. To deduce invariants, its source code is analyzed (specially data and control flows). On one hand, invariants generated this way are always correct. But, on the other hand, its number and quality is limited due to the inner limitations of the formal machinery which analyzes the code, specially in unusual languages like WS-BPEL.

Conversely, a dynamic invariant generator (Ernst et al., 2001) is a system that reports likely program invariants observed on a set of execution log files. It includes formal machinery to analyze the information in the logs about the values held by variables at different program points, such as the entry and exit points for functions or loops.

The process to generate dynamic invariants is divided into three main steps:

1. **An instrumentation step** where the original program is set up so that, during the later execution step, it generates the execution log files. This step is called instrumentation step because the usual way to do it is by adding, at the desired program points, logging instructions. These instructions write to a file the name and value of the variables that we want to observe at those points and otherwise have no effect on the control and data flow of the process. Sometimes it is also necessary to modify the environment where the program is going to be executed.

2. **An execution step** in which the instrumented program will be executed under a test suite. During each test case an execution log is generated with all the necessary data and program flow information for later processing.

3. **An analysis step** where formal methods techniques are applied to obtain invariants of the variables logged at the different program points.

Thus, the dynamic generation of invariants does not analyze the code, but a set of samples of the values held by variables in certain points of the program. Wrong invariants do not necessarily mean bugs in the tested program, but rather they might come from an incomplete test suite. If the input \( x \) is a signed integer and we only used positive values as test inputs, we will probably obtain the invariant \( x > 0 \) at some program point. Upon inspection, we would notice that invariant and improve our test suite to including cases with \( x < 0 \).

### 3.3 Dynamic Invariant Generation in WS-BPEL Compositions

We consider the dynamic generation of invariants to be a suitable technique to support WS-BPEL composition white-box testing. If we use a good test suite, all of the complex internal logic of our BPEL composition (compensation, dynamic discovery of services, etc.) will be reflected in the log files of the different
executions, and the generator will infer significant invariants.

Generally, due to its dynamic nature, the more logs we provide the generator, the better results it will produce. In case we obtain seemingly false invariants in a first run, we will be able to certify if they were due to an incomplete test suite or to actual bugs in a second run with an improved test suite including additional suitable test cases.

Another important benefit is that all the information in the logs is collected from direct execution of the composition code, using no intermediate language of any sort. This way we avoid errors that could arise in any translation of the WS-BPEL code or the simulation of the real-world environment, that is, the WS-BPEL engine and invoked services.

An important problem to solve is that usually, all external services will not be available for testing, due to access restrictions, reliability issues or resource constraints. Or it could also just be that we wanted to define several what-if scenarios with specific responses from several external WS. Thus, we will also have to allow for replacing some external services with mockups, that is, dummy services which will reply our requests with predefined messages.

4 PROPOSED ARCHITECTURE

An outline of our proposed architecture for a WS-BPEL dynamic invariant generator is shown in the figure 1.

We are using a classical pipeline-based architecture which has, in general terms, three coarse-grained steps corresponding to the three general steps of the dynamic invariant generation process which we described above.

We detail further their WS-BPEL specific issues below.

4.1 Instrumentation Step

In this is the step where we take the original program and perform the necessary changes on it in order to produce the information that the invariant generator needs.

In our case, we will take the original WS-BPEL process composition specification with its dependencies and automatically instrument it. If necessary we will create any additional files needed for its execution in the specific WS-BPEL engine which we will be using.

Additional logic to generate the logs that we need can be added in three ways:

1. Firstly, we could modify the execution environment itself. We could use an existing open-source WS-BPEL engine and modify it in order to produce the log files that we need for any process executed in it.

   The degree of effort involved would depend on the logging capabilities already implemented in the engine. Most engines include facilities to track process execution flow, but tracking variable values is not widely implemented. Of course, modifying the code would take a considerable effort.

2. Secondly, we could modify the source code of the WS-BPEL composition to be tested, adding calls to certain logging instructions at the desired program points. These instructions would not change the behavior of the process, being limited to transparently inspect and log variable values.

   In a similar way to the previous method, we might be able to use any existing engine-specific WS-BPEL logging extension. It could append messages to a log file, access a database or invoke an external logging web service, for instance.

   Using this approach we would have to instrument the two different languages used by the WS-BPEL standard: the WS-BPEL language itself, which is XML-based, and the XPath language that is used to construct complex expressions for conditions, assignments, and so on.

3. And finally, we could implement our own logging XPath extension functions.

   These new functions would be called from an instrumented version of the original WS-BPEL composition source code, and included in external modules which are reasonably easy to automatically plug into most WS-BPEL engines. This is a hybrid approach, as both the composition and the engine need to be modified.

   It is quite likely that using both internal (that is, inside the WS-BPEL engine) and external modifications (using new XPath extension functions) will allow us to obtain more detailed logs with less effort than any of the previous approaches.

We also take into account the fact that there are many WS-BPEL engines currently available. Engine-specific files for the deployment of the composition under the selected engine may have to be generated automatically on the fly, to abstract the user from the technical details involved.

4.2 Execution Step

In this step, we will take the previously instrumented program and run it under a test suite to obtain the logs
and other information that we need for the next step. Specifically, we will take our instrumented WS-BPEL composition and run it under each test case detailed in the test suite specification. Each of them will define the initial input message that will cause a new instance of the WS-BPEL process to be started, as well
as the outputs of the external services required by the WS-BPEL process that we wish to model as mockups.

The ability to model none, some or even all of the external services as mockups will enable us to obtain invariants reflecting different situations. On one hand, if we do not use mockups at all, but only invoke actual services, we will be studying the complete WS-BPEL composition in the real-world environment. On the other, if every external service is replaced with a mockup with predefined responses, we will be focusing on the internal logic of the composition itself and how it behaves in certain scenarios. We can also settle for a middle point in a hybrid approach.

After this step is done, every test case will have generated its own execution log, which we will pass on to the next step. We will need roughly two components to make this possible:

- A WS-BPEL engine for running the process itself, which will invoke the external services as needed. Making the engine use a mockup for an external service could be achieved in basically two ways: by modifying the service address included in the WSDL source files, or by creating (or modifying) the engine-specific files with the proper values.

- A WS-BPEL unit test library which will deploy and act both as a client, invoking our composition with the desired parameters, and as the external mockup services for the WS-BPEL process. These services will behave according to the external test case specification described above.

This unit test library has to be quite more complex than similar libraries for other languages. It can be divided once more into the following subcomponents:

- A director which will prepare and monitor the execution of the whole test suite according to each test case specification. Ideally, it should also be able to deploy and undeploy the WS-BPEL process from our selected engine.

- A mockup server, properly set up by the director, which will handle incoming requests and act as the external mockup services required by the WS-BPEL process that we choose to model. Mockups have no internal logic of their own, being limited to either replying with a predefined XML SOAP message or failing as indicated in the test case specification.

There are several lightweight Java-based web servers available for this role, but, if necessary, we believe it would also be feasible to develop it from scratch, being a simple URL → SOAP message matching system.

4.3 Analysis Step

After all the test cases have been executed and logs have been collected, it would seem at first glance to be a matter of just handing them to our invariant generator.

However, it will not be so simple in most cases, because the invariant generator could require additional information about the data to analyze to work properly. All information would have to be reformatted according to the input format expected by the invariant generator. This reformattting could range from a simple translation to a thorough transformation of the XML data structures to those available to the invariant generator.

The invariant generator may even accept not only logs, but also a list of constraints already known, and which thus do not need to be generated again as invariants. This would reduce its output size and make it easier to understand for its users. To generate this constraint list, we would have to analyze the XML Schema files contained in the WS-BPEL process and all of its dependencies.

All these obstacles can be overcome through a preprocessor. It could even call the invariant generator with the generated files to finally obtain the invariants of the WS-BPEL composition.

Depending on the number and complexity of the invariants produced by the generator, we could even need to pass them later to a simplifier. It is a program based on formal methods that receives a set of invariants and removes those logically inferred by others.

4.4 Example

We comment briefly an example of the invariants we could infer in the classical WS-BPEL example of the Loan Approval Service included in the WS-BPEL 2.0 specification (OASIS, 2007).

This WS-BPEL composition receives loan requests from costumers. Each request includes the amount and certain personal information. The amount and certain personal information. The WS-BPEL composition simply notifies the costumer whether his loan request has been approved or rejected. The approval of the loan is based on the amount requested and the risk that a risk assessment determines for the costumer according to his personal information. If the amount is below $10,000 and the risk assessment WS considers the applicant a low-risk costumer the loan is automatically approved.

In case the amount is below the threshold but risk is considered medium or high, the composition invokes an external loan approval WS, and its answer is passed to the costumer as the response of the com-
position. Finally, in case the requested amount is over the threshold no risk checking is done, and the answer of the composition is also that of the external loan approval WS.

Our architecture could infer the following invariants for this example, it were backed by an exhaustive and high-quality test suite:

\[ \text{request.amount} < 10000 \land \text{risk.level} = 'low' \implies \text{approval.accept} = 'yes' \]

\[ \text{request.amount} < 10000 \land \text{risk.level} \neq 'low' \implies \text{approval.accept} = \text{response(approve)} \]

\[ \text{request.amount} \geq 10000 \implies \text{approval.accept} = \text{response(approve)} \]

invoke(customer) = approval

We can clearly see that the system could infer that approval depends on the amount requested and the response provided by the approver when invoked. The system could also detect that the response we provide the costumer is always the value of the variable approval.

Of course, to get results this fine-grained we will need a good test suite. For our example, the test suite would have to contain test cases with amounts over and under the threshold (including the limit values \$9,999, \$10,000 and \$10,001). Personal information causing the risk assessment WS answers both affirmatively and negatively must also be provided, specially for those under the threshold. At any rate, as discussed before, in case we do not obtain the desired invariants in the first run, we could extend the test suite with more test cases refining the invariants obtained. This way, in the next run, we would obtain more accurate invariants.

5 RELATED WORK

In this section we present some related works. There is a wide variety of topics related to our architecture, mainly dynamic invariant generation, WS composition testing and test case generation:

An interesting proposal to use dynamically invariants for WS quality testing is (SeCSE, 2007). It collects several invocations and replies from a WS-BPEL composition to an external WS and dynamically generates likely invariants to check its Service Level Agreement. Therefore, it constitutes a black-box testing technique. In contrast, our architecture follows a white-box testing approach, oriented to the generation of the invariants from the internal logic of a WS-BPEL composition.

The relation between the test cases used for dynamically generating invariants and the quality of the invariants derived is studied in (Gupta, 2003). Augmenting a test suite with suitable test cases can be an interesting way to increase the accuracy of the invariants inferred by our architecture.

Dynamo (Baresi and Guinea, 2005) is a proxy-based system to monitor if a WS-BPEL composition holds several restrictions during its execution. We think it might be useful as a way to check if the invariants obtained from our architecture hold while it is running in a real-world environment.

Test cases for a WS-BPEL composition are automatically generated in (Zheng et al., 2007) according to state and transition coverage criteria. We could assure the quality of the invariants generated by using them as inputs for our architecture.

6 CONCLUSIONS AND FUTURE WORK

The fact that WS are the future of e-Business is already a given, thanks to their platform independence and the abstractions that they provide. The need to orchestrate them to provide more advanced services suiting costumer’s requirement has been satisfied through the WS-BPEL 2.0 standard.

However, WS-BPEL compositions are difficult to test, since traditional white-box testing techniques are difficult to apply to them. This is because of the unusual mix of features present in WS-BPEL, such as concurrency support, event handling or fault compensation. We have showed how dynamically generated likely invariants backed by a good test suite can become a suitable and successful help to solve these difficulties, thanks to their being based on actual execution logs.

In this work, we have proposed a pipelined architecture for dynamically generating invariants from a WS-BPEL composition. Requirements on every of the components have been identified, leaving as our next future work finding suitable systems for each one.

Once the architecture is completely implemented we will perform an experimental evaluation of the framework under several compositions. This way we will test its reliability and evaluate its results through metrics such as quality of the invariants generated or time
taken to infer them.

Looking further ahead, we will later study the relation between the quality of the invariants generated and the test case suite used to infer them. For this we can use different WS-BPEL compositions with their specifications and different test suites providing certain coverage criteria (branch coverage, statement coverage, . . . ) of them.

Finally, we could use the invariants generated by our proposal to support WS-BPEL white-box testing and check if it improves its results.

ACKNOWLEDGEMENTS

This work has been financed by the Programa Nacional de I+D+I of the Spanish Ministerio de Educaciön y Ciencia and FEDER funds through SOAQSim project (TIN2007-67843-C06-04).

REFERENCES


A NEW REINFORCEMENT SCHEME FOR STOCHASTIC LEARNING AUTOMATA

Application to Automatic Control

Florin Stoica, Emil M. Popa
Computer Science Department, “Lucian Blaga” University, Str. Dr. Ion Ratiu 5-7, Sibiu, Romania
florin.stoica@ulbsibiu.ro, emilm.popa@ulbsibiu.ro

Iulian Pah
Department of Sociology, “Babes-Bolyai” University, Bd.21 decembrie 1989, no.128-130, Cluj-Napoca, Romania
i_pah@yahoo.com

Keywords: Stochastic Learning Automata, Reinforcement Learning, Intelligent Vehicle Control, agents.

Abstract: A Learning Automaton is a learning entity that learns the optimal action to use from its set of possible actions. It does this by performing actions toward an environment and analyzes the resulting response. The response, being both good and bad, results in behaviour change to the automaton (the automaton will learn based on this response). This behaviour change is often called reinforcement algorithm. The term stochastic emphasizes the adaptive nature of the automaton: environment output is stochastically related to the automaton action. The reinforcement scheme presented in this paper is shown to satisfy all necessary and sufficient conditions for absolute expediency for a stationary environment. An automaton using this scheme is guaranteed to „do better” at every time step than at the previous step. Some simulation results are presented, which prove that our algorithm converges to a solution faster than one previously defined in (Ünsal, 1999). Using Stochastic Learning Automata techniques, we introduce a decision/control method for intelligent vehicles, in infrastructure managed architecture. The aim is to design an automata system that can learn the best possible action based on the data received from on-board sensors or from the localization system of highway infrastructure. A multi-agent approach is used for effective implementation. Each vehicle has associated a “driver” agent, hosted on a JADE platform.

1 INTRODUCTION

The past and present research on vehicle control emphasizes the importance of new methodologies in order to obtain stable longitudinal and lateral control. In this paper, we consider stochastic learning automata as intelligent controller within our model for an Intelligent Vehicle Control System.

An automaton is a machine or control mechanism designed to automatically follow a predetermined sequence of operations or respond to encoded instructions. The term stochastic emphasizes the adaptive nature of the automaton we describe here. The automaton described here does not follow predetermined rules, but adapts to changes in its environment. This adaptation is the result of the learning process (Barto, 2003). Learning is defined as any permanent change in behavior as a result of past experience, and a learning system should therefore have the ability to improve its behavior with time, toward a final goal.

The stochastic automaton attempts a solution of the problem without any information on the optimal action (initially, equal probabilities are attached to all the actions). One action is selected at random, the response from the environment is observed, action probabilities are updated based on that response, and the procedure is repeated. A stochastic automaton acting as described to improve its performance is called a learning automaton. The algorithm that guarantees the desired learning process is called a reinforcement scheme (Moody, 2004).

Mathematically, the environment is defined by a triple \( \{ \alpha, c, \beta \} \) where \( \alpha = \{ \alpha_1, \alpha_2, ..., \alpha_c \} \) represents a finite set of actions being the input to the environment, \( \beta = \{ \beta_1, \beta_2 \} \) represents a binary response set, and \( c = \{ c_1, c_2, ..., c_c \} \) is a set of penalty
probabilities, where \( c_i \) is the probability that action \( \alpha_i \) will result in an unfavorable response. Given that \( \beta(n) = 0 \) is a favorable outcome and \( \beta(n) = 1 \) is an unfavorable outcome at time instant \( n \ (n = 0, 1, 2, \ldots) \), the element \( c_i \) of \( c \) is defined mathematically by:

\[
c_i = P(\beta(n) = 1 | \alpha(n) = \alpha_i) \quad i = 1, 2, \ldots, r
\]

The environment can further be split up in two types, stationary and nonstationary. In a stationary environment the penalty probabilities will never change. In a nonstationary environment the penalties will change over time.

In order to describe the reinforcement schemes, is defined \( p(n) \), a vector of action probabilities:

\[
p_i(n) = P(\alpha(n) = \alpha_i), \quad i = 1, r
\]

Updating action probabilities can be represented as follows:

\[
p(n+1) = T[p(n), \alpha(n), \beta(n)]
\]

where \( T \) is a mapping. This formula says the next action probability \( p(n+1) \) is updated based on the current probability \( p(n) \), the input from the environment and the resulting action. If \( p(n+1) \) is a linear function of \( p(n) \), the reinforcement scheme is said to be linear; otherwise it is termed nonlinear.

## 2 REINFORCEMENT SCHEMES

### 2.1 Performance Evaluation

Consider a stationary random environment with penalty probabilities \( \{c_1, c_2, \ldots, c_r\} \) defined above.

We define a quantity \( M(n) \) as the average penalty for a given action probability vector:

\[
M(n) = \sum_i c_i p_i(n)
\]

An automaton is absolutely expedient if the expected value of the average penalty at one iteration step is less than it was at the previous step for all steps: \( M(n+1) < M(n) \) for all \( n \) (Rivero, 2003).

The algorithm which we will present in this paper is derived from a nonlinear absolutely expedient reinforcement scheme presented by (Ünsal, 1999).

### 2.2 Absolutely Expedient Reinforcement Schemes

The reinforcement scheme is the basis of the learning process for learning automata. The general solution for absolutely expedient schemes was found by (Lakshmivarahan, 1973).

A learning automaton may send its action to multiple environments at the same time. In that case, the action of the automaton results in a vector of responses from environments (or "teachers"). In a stationary \( N \)-teacher environment, if an automaton produced the action \( \alpha_i \) and the environment responses are \( \beta_j \ j = 1, \ldots, N \) at time instant \( n \), then the vector of action probabilities \( p(n) \) is updated as follows (Ünsal, 1999):

\[
p_i(n+1) = p_i(n) + \left[ \frac{1}{N} \sum_{j=1}^{N} \beta_j \right] \cdot \phi_i(p(n)) - \left[ \frac{1}{N} \sum_{j=1}^{N} \beta_j \right] \cdot \psi_i(p(n))
\]

\[
p_j(n+1) = p_j(n) - \left[ \frac{1}{N} \sum_{j=1}^{N} \beta_j \right] \cdot \phi_j(p(n)) + \left[ \frac{1}{N} \sum_{j=1}^{N} \beta_j \right] \cdot \psi_j(p(n))
\]

for all \( j \neq i \) where the functions \( \phi_i \) and \( \psi_i \) satisfy the following conditions:

\[
\frac{\phi_i(p(n))}{p_i(n)} = \ldots = \frac{\phi_i(p(n))}{p_i(n)} = \lambda(p(n))
\]

\[
\frac{\psi_i(p(n))}{p_i(n)} = \ldots = \frac{\psi_i(p(n))}{p_i(n)} = \mu(p(n))
\]

\[
p_i(n) + \sum_{j \neq i} \phi_j(p(n)) > 0
\]

\[
p_i(n) - \sum_{j \neq i} \psi_j(p(n)) < 1
\]

\[
p_j(n) + \psi_j(p(n)) > 0
\]

\[
p_j(n) - \phi_j(p(n)) < 1
\]

for all \( j \in \{1, \ldots, r\} \setminus \{i\} \).

The conditions (3)-(6) ensure that \( 0 < p_k < 1, \ k = 1, r \) (Stoica, 2007).

**Theorem.** If the functions \( \lambda(p(n)) \) and \( \mu(p(n)) \) satisfy the following conditions:
\( \lambda(p(n)) \leq 0 \\
\mu(p(n)) \leq 0 \\
\lambda(p(n)) + \mu(p(n)) < 0 \)  
then the automaton with the reinforcement scheme in (1) is absolutely expedient in a stationary environment.

The proof of this theorem can be found in (Baba, 1984).

3 A NEW NONLINEAR REINFORCEMENT SCHEME

Because the above theorem is also valid for a single-teacher model, we can define a single environment response that is a function \( f \) of many teacher outputs.

Thus, we can update the above algorithm as follows:

\[
p_i(n+1) = p_i(n) + f \left( -\theta \delta H(n) \right) \left[ 1 - p_i(n) \right] - (1 - f) \left( -\theta \right) \left[ 1 - p_i(n) \right] \\
p_j(n+1) = p_j(n) - f \left( -\theta \delta H(n) \right) \left[ 1 - p_j(n) \right] + (1 - f) \left( -\theta \right) \left[ 1 - p_j(n) \right]
\]

for all \( j \neq i \), i.e.:

\[
\psi_i(p(n)) = -\theta \delta H(n) \left[ 1 - p_i(n) \right] \\
\phi_j(p(n)) = -\theta \delta H(n) \left[ 1 - p_j(n) \right]
\]

where learning parameters \( \theta \) and \( \delta \) are real values which satisfy:

\( 0 < \theta < 1 \) and \( 0 < \theta \delta < 1 \).

The function \( H \) is defined as follows:

\[
H(n) = \min\left\{ \begin{array}{c}
\frac{p_i(n)}{\theta \delta \left( 1 - p_i(n) \right) - \varepsilon} \\
\frac{1 - p_i(n)}{\theta \delta \left( 1 - p_i(n) \right) - \varepsilon}
\end{array} \right\} \left( \lambda \left( \frac{1}{i} \right) \right)
\]

Parameter \( \varepsilon \) is an arbitrarily small positive real number.

Our reinforcement scheme differs from the one given in (Ünsal, 1999) by the definition of these two functions: \( H \) and \( \phi_k \).

The proof that all the conditions of the reinforcement scheme (1) and theorem (7) are satisfied can be found in (Stoica, 2007).

In conclusion, we state the algorithm given in equations (8) is absolutely expedient in a stationary environment.

4 EXPERIMENTAL RESULTS

4.1 Problem Formulation

To show that our algorithm converges to a solution faster than the one given in (Ünsal, 1999), let us consider a simple example. Figure 1 illustrates a grid world in which a robot navigates. Shaded cells represent barriers.

Figure 1: A grid world for robot navigation.

The current position of the robot is marked by a circle. Navigation is done using four actions \( \alpha = \{N, S, E, W\} \), the actions denoting the four possible movements along the coordinate directions.

Because in given situation there is a single optimal action, we stop the execution when the probability of the optimal action reaches a certain value (0.9999).

4.2 Comparative Results

We compared two reinforcement schemes using these four actions and two different initial conditions.

Table 1: Convergence rates for a single optimal action of a 4-action automaton (200 runs for each parameter set).

<table>
<thead>
<tr>
<th>Average number of steps to reach ( p_{\text{opt}}=0.9999 )</th>
<th>4 actions with ( p_{\text{opt}}(0)=1/4 ), ( i=1/4 )</th>
<th>4 actions with ( p_{\text{opt}}(0)=0.0005 ), ( i=0.0005/3 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \theta )</td>
<td>( \delta )</td>
<td>Ünsal's Alg.</td>
</tr>
<tr>
<td>0.01</td>
<td>25</td>
<td>644.84</td>
</tr>
<tr>
<td>0.05</td>
<td>5</td>
<td>136.99</td>
</tr>
<tr>
<td>0.1</td>
<td>10</td>
<td>24.74</td>
</tr>
<tr>
<td>2.5</td>
<td>5</td>
<td>74.05</td>
</tr>
<tr>
<td>5</td>
<td>2.5</td>
<td>59.48</td>
</tr>
<tr>
<td>5</td>
<td>23.05</td>
<td>19.51</td>
</tr>
</tbody>
</table>

The data shown in Table 1 are the results of two different initial conditions where in first case all
probabilities are initially the same and in second case the optimal action initially has a small probability value (0.0005), with only one action receiving reward (i.e., optimal action).

Comparing values from corresponding columns, we conclude that our algorithm converges to a solution faster than the one given in (Ünsal, 1999).

5 USING STOCHASTIC LEARNING AUTOMATA FOR INTELLIGENT VEHICLE CONTROL

The task of creating intelligent systems that we can rely on is not trivial. In this section, we present a method for intelligent vehicle control, having as theoretical background Stochastic Learning Automata. We visualize the planning layer of an intelligent vehicle as an automaton (or automata group) in a nonstationary environment. We attempt to find a way to make intelligent decisions here, having as objectives conformance with traffic parameters imposed by the highway infrastructure (management system and global control), and improved safety by minimizing crash risk.

The aim here is to design an automata system that can learn the best possible action based on the data received from on-board sensors, or from roadside-to-vehicle communications. For our model, we assume that an intelligent vehicle is capable of two sets of lateral and longitudinal actions. Lateral actions are LEFT (shift to left lane), RIGHT (shift to right lane) and LINE_OK (stay in current lane). Longitudinal actions are ACC (accelerate), DEC (decelerate) and SPEED_OK (keep current speed). An autonomous vehicle must be able to “sense” the environment around itself. Therefore, we assume that there are four different sensors modules on board the vehicle (the headway module, two side modules and a speed module), in order to detect the presence of a vehicle traveling in front of the vehicle or in the immediately adjacent lane and to know the current speed of the vehicle.

These sensor modules evaluate the information received from the on-board sensors or from the highway infrastructure in the light of the current automata actions, and send a response to the automata. Our basic model for planning and coordination of lane changing and speed control is shown in Figure 2.

Figure 2: The model of the Intelligent Vehicle Control System.

The response from physical environment is a combination of outputs from the sensor modules. Because an input parameter for the decision blocks is the action chosen by the stochastic automaton, it is necessary to use two distinct functions $F_1$ and $F_2$ for mapping the outputs of decision blocks in inputs for the two learning automata, namely the longitudinal automaton and respectively the lateral automaton.

After updating the action probability vectors in both learning automata, using the nonlinear reinforcement scheme presented in section 3, the outputs from stochastic automata are transmitted to the regulation layer. The regulation layer handles the actions received from the two automata in a distinct manner, using for each of them a regulation
buffer. If an action received was rewarded, it will be introduced in the regulation buffer of the corresponding automaton, else in buffer will be introduced a certain value which denotes a penalized action by the physical environment. The regulation layer does not carry out the action chosen immediately; instead, it carries out an action only if it is recommended \( k \) times consecutively by the automaton, where \( k \) is the length of the regulation buffer. After an action is executed, the action probability vector is initialized to \( \frac{1}{r} \), where \( r \) is the number of actions. When an action is executed, regulation buffer is initialized also.

6 SENSOR MODULES

The four teacher modules mentioned above are decision blocks that calculate the response (reward/penalty), based on the last chosen action of automaton. Table 2 describes the output of decision blocks for side sensors.

Table 2: Outputs from the Left/Right Sensor Module.

<table>
<thead>
<tr>
<th>Actions</th>
<th>Left/Right Sensor Module</th>
<th>No vehicle in sensor range or no adjacent lane</th>
</tr>
</thead>
<tbody>
<tr>
<td>LINE_OK</td>
<td>0/0</td>
<td>0/0</td>
</tr>
<tr>
<td>LEFT</td>
<td>1/0</td>
<td>0/0</td>
</tr>
<tr>
<td>RIGHT</td>
<td>0/1</td>
<td>0/0</td>
</tr>
</tbody>
</table>

Table 3: Outputs from the Headway Module.

<table>
<thead>
<tr>
<th>Actions</th>
<th>Headway Sensor Module</th>
<th>No vehicle in range (at a close frontal distance)</th>
</tr>
</thead>
<tbody>
<tr>
<td>LINE_OK</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>LEFT</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>RIGHT</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>SPEED_OK</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>ACC</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>DEC</td>
<td>0*</td>
<td>0</td>
</tr>
</tbody>
</table>

As seen in Table 2, a penalty response is received from the left sensor module when the action is LEFT and there is a vehicle in the left or the vehicle is already traveling on the leftmost lane. There is a similar situation for the right sensor module.

The Headway (Frontal) Module is defined as shown in Table 3. If there is a vehicle at a close distance (< admissible distance), a penalty response is sent to the automaton for actions LINE_OK, SPEED_OK and ACC. All other actions (LEFT, RIGHT, DEC) are encouraged, because they may serve to avoid a collision.

The Speed Module compares the actual speed with the desired speed, and based on the action choosed send a feedback to the longitudinal automaton.

Table 4: Outputs from the Speed Module.

<table>
<thead>
<tr>
<th>Actions</th>
<th>Speed Sensor Module</th>
</tr>
</thead>
<tbody>
<tr>
<td>SPEED_OK</td>
<td>1</td>
</tr>
<tr>
<td>ACC</td>
<td>0</td>
</tr>
<tr>
<td>DEC</td>
<td>1</td>
</tr>
</tbody>
</table>

The reward response indicated by 0* (from the Headway Sensor Module) is different than the normal reward response, indicated by 0: this reward response has a higher priority and must override a possible penalty from other modules.

7 A MULTI-AGENT SYSTEM FOR INTELLIGENT VEHICLE CONTROL

In this section is described an implementation of a simulator for the Intelligent Vehicle Control System, in a multi-agent approach. The entire system was implemented in Java, and is based on JADE platform (Bigus, 2001).

In figure 3 is showed the class diagram of the simulator. Each vehicle has associated a JADE agent (DriverAgent), responsible for the intelligent control. “Driving” means a continuous learning process, sustained by the two stochastic learning automata, namely the longitudinal automaton and respectively the lateral automaton.

The response of the physical environment is a combination of the outputs of all four sensor modules. The implementation of this combination for each automaton (longitudinal respectively lateral) is showed in figure 4 (the value 0* was substituted by 2).
public double reward(int action){
    int combine;
    combine=Math.max(speedModule(action),
                     frontModule(action));
    if (combine == 2) combine = 0;
    return combine;
}

public double reward(int action){
    int combine;
    combine=Math.max(
                     leftRightModule(action),
                     frontModule(action));
    return combine;
}

Figure 4: The physical environment response.

8 CONCLUSIONS

Reinforcement learning has attracted rapidly increasing interest in the machine learning and artificial intelligence communities. Its promise is beguiling - a way of programming agents by reward and punishment without needing to specify how the task (i.e., behavior) is to be achieved. Reinforcement learning allows, at least in principle, to bypass the problems of building an explicit model of the behavior to be synthesized and its counterpart, a meaningful learning base (supervised learning).

The reinforcement scheme presented in this paper satisfies all necessary and sufficient conditions for absolute expediency in a stationary environment and the nonlinear algorithm based on this scheme is found to converge to the "optimal" action faster than nonlinear schemes previously defined in (Ünsal, 1999).

Using this new reinforcement scheme was developed a simulator for an Intelligent Vehicle Control System, in a multi-agent approach. The entire system was implemented in Java, and is based on JADE platform.

REFERENCES

BUSINESS INTELLIGENCE THROUGH REAL-TIME TRACKING

Using a Location System Towards Behaviour Pattern Extraction

Pedro Abreu, Vasco Vinhas
FEUP - Faculty of Engineering of the University of Porto, R. Campo Alegre 1021, Porto, Portugal
DEI - Department of Informatics Engineering, R. Campo Alegre 1021, Porto, Portugal
LIACC - Artificial Intelligence and Computer Science Laboratory, R. Campo Alegre 1021, Porto, Portugal
pha@fe.up.pt, vvm@fe.up.pt

Pedro Mendes
FEUP - Faculty of Engineering of the University of Porto, R. Campo Alegre 1021, Porto, Portugal
DEI - Department of Informatics Engineering, R. Campo Alegre 1021, Porto, Portugal
INESC Porto - Institute for Systems and Computer Engineering of Porto, Rua Dr. Roberto Frias, 378, Porto, Portugal
pmcm@fe.up.pt

Keywords: Location System, Real-Time, Knowledge Extraction, Business Intelligence.

Abstract: Nowadays, tracking systems constitute an important knowledge support in order to compute important measurements in companies processes efficiency. As consequence of that, this project proposes a methodology and an application, based on a tracking system to obtain, by automatic means, dynamic location data on items. This solution assumes that the client carries or drives an item of some kind. In each item there is an identifying tag attached and hidden in order to make the item at hand detectable by all the sensors that are scattered around the area. Because of the fact that the tag is light and hidden and also has no information regarding the specific person/agent this process is completely transparent to the client or robot that is being implicitly tracked. This system produces real-time shop floor visualization maps with intelligible data on online item localization; individual item complete path routes; online and historical population density rates and path routes concentration; and also item vision enabled concentration maps as emulation for item omnidirectional vision considering occlusions. This proposed system might be useful in many different areas, for instance in a traditional retail environment tracing clients through a commercial area or enabling item tracking and route analysis in a hospital.

1 INTRODUCTION

Detecting behavioral patterns is a challenging task that marketing and distribution companies face. The issue has been addressed through the past years on several perspectives like deterministic psychology (Luce, 1999; Choustova, 2007). However due to their active consciousness, human beings are extremely unpredictable and so these methods failed to provide any accurate data that could be used for industrial purposes. Having these approaches been unsuccessful, statistical inference with large data sets is still one of the most powerful tools available.

Nowadays, tracking systems can represent a powerful tool to support monitoring activity. With these systems, some performance measurements on company process efficiency can be obtained regardless of the specific tracked target. This research work presents a methodology and a tool, based on a tracking system to obtain by automatic means, movements data on these elements.

The presented solution assumes that the client carries or drives an item of some kind, inside the space. It is also assumed that these items are outside the entry of the traceable space, although already inside of its admissible space. In each of the items there is an identifying tag attached and hidden in order to make the item at hand detectable by all the radars that are scattered around the area. By detecting it, the client or robot is being implicitly tracked in a completely transparent way for him/it since the tag is light and hidden and also has no information regarding the specific person/agent.

Several benefits can be withdrawn from using a system such as this. Instantaneously one could use it to monitor the traceable area in a more effective way...
than looking at dozens of screens with images from security cameras. Live monitoring of the elements’ positions on a specific area allows managers to identify congested sectors. It also allows him to identify hot and cold zones which may be related to the interests points at hand or, for instance, to a local security issue or other type of event such a medical emergence.

Regarding long term data analysis the advantages of using such a system are several and may be more than the ones presented next. By analyzing all the paths taken by the elements, it is possible to obtain the hot zones on any time frame and thus evaluate the success rate of a given promotional campaign and among other things. Erroneous and random movements may also be correlated with a security issue, and thus this system could also represent an interesting addition to conventional security systems. The results later obtained would clearly point out the success of the layout redesign. Of course these last measures imply correlating client positions with goods bought by him.

The paper is structured as follows: section 2 describes the current state of the art regarding the several areas of knowledge involved in the development of the system described in this paper. Section 3 describes the tool that simulates the environment as well as its architecture and core functionalities. Section 4 discusses the results obtained so far and the next section concludes the paper by summarizing the focus of this research work and pointing out future lines of discussion.

2 STATE-OF-THE-ART

Nowadays, tracking systems represent an important research area as their applications are transversal to several areas of knowledge (e.g. computer science, medicine, simulation, robotics as well as industrial tracks). In the past few years, technology has evolved in order to provide more accurate measurements. In the robotics area, for better modeling the world, it is extremely relevant to accurately process the signals received by the multiple sensors involved. Locating objects of the real world to the modeled one is a critical task for the appliance of the navigation algorithms and methodologies. Following these advances the work published by Hyunwoong Park (Park H., 2006) presents a new kind of sensor system with multiple rotating range sensors. Such system allows a robot to guide itself on a priori unknown world. On the other hand these tracking systems also find interesting applications on scenarios where the context environment is already known. Regarding this last system, locating elements assumes a crucial role. To achieve this goal, several technologies have been used. By doing a brief comparison, it is observable that all of them have their strengths and flaws concerning characteristics like the cost in terms of initial investment and maintenance. There are others related to environment specificities. Among these last, other parameters such as coverable area, tracking detection errors and occlusion problems should also be considered.

One of the most effective technologies is also one of the most expensive ones and concerns detection of thermal signatures. This technique is appropriate to living organisms which emit particular heat waves. One particular application of this technology is the monitoring of the fauna in the ocean (Raizer, 2003). Another interesting technology is Bluetooth because most modern mobile equipments are prepared to send a receive data though this protocol. Although the initial investment is low the coverable area is not very wide and battery consumption is high, in relative terms (Jappinen P., 2007).

The cheapest solution is infrared based. Even though its price attractiveness, infrared systems tend to fail on most real environments because the signal is unable to reach the target if there is an opaque object between the receiver and the target (Krotsosky J., 2007).

Two of the most emergent technologies for tracking are RFID and Wi-Fi based. The first one still lacks standardization which is somehow reflected in the pricing of both receivers and transmitters. It is based on high frequency radio waves having the detectable tags a passive or an active response. Passive tags are only detectable on a 13 meter radius and are used for instance on the new USA passports. Active tags, alternatively, are detectable on a much wider range but are more expensive (around 400%) essentially because the tags require an independent power supply (Chao C., 2007). Wi-Fi may also be considered as a tracking technology. This approach is mainly used for creating wireless computer networks but in this case the involved tracking only requires the usage of the low level protocols. This type of solution is interesting because it makes possible using existent computer networks for other proposes and takes advantages from possible simple detection with at least one access point. With only one access point the system’s precision may not be very high but there is no need for triangulation. Occlusion problems and signals losses, with the use of this technique will be reduced to a residual level in both open spaces and indoors – considering that indoor spaces do not have significant metal structures within the walls) (Minghkhwan, 2006).

Another area that suffered several developments
in the past few years concerns world modeling. In this research area there are relevant research topics, most of them related to computer graphics. Most of the current advances focus on three dimensional (3D) worlds. In this scope, the evolution on computer graphics is the most notorious. Nowadays, simple systems are able to represent complex 3D world including high resolution textures, detail animation (Vazquez, 2007) and weather condition (Grudzinski, 2007). It is even possible to recreate a 3D world from textual specifications (Moura J., 2004). On top of 3D world, many algorithms are applicable in order to optimize rendering performance and obtaining world data such as visible objects of a certain point.

These algorithms are too complex for most real-time tracking systems and therefore for this research work the world is assumed to be a two and a half dimension one (2,5D). In this scenario, a map is represented considering a bird’s eye view and assuming that the height of the objects has no maximum value. In these conceptions it is simpler to obtain the set of potentially visible objects from a certain point using a portal culling algorithms; that consider walls as complete occluders and assigns a vision probability to each region in the map (Pires, 2001). Such is achieved by dividing the indoor space into separate sectors and then portals which represent the breaches between the sectors. By drawing cones that connect the observer point and both extremities of a given portal, one can obtain the areas where all the objects are potentially visible. Some caution is required when performing such an operation, since the lines that represent the vision cone cannot intersect the ones representing zone divisions. It is also relevant to state that one must assume the observer’s vision direction is the center of a given "vision cone".

3 PROJECT DESCRIPTION

The project description is divided of three distinct subsections. In the first denominated as Project Architecture a description of the system’s architecture is depicted. In the Real Time Tracking Visualization and Concentration Maps subsection the principles of the system features is explained and in the last section, Client Vision Module the used vision algorithm is exposed.

3.1 Project Architecture

This research work proposes a decentralized architecture and prototype tool following that same principle as detailed in Figure 1.

![Figure 1: System’s Architecture.](image-url)

In the first stage, using a location system that might be instantiated in a RFID or Wi-Fi based solution, covering an area of for instance a large open space or building, with a maximum error of one meter, movement data is collected regarding the tagged elements present in the given specific floor. In order to gather the location data one agent has been developed to collect all the positions. This action has a given periodicity and is dependent on the location engine. Typically, this collected data is guaranteed to be obtained every second at most for every single tag, although this figure might be decreased depending on the number of simultaneously trackable objects.

After the collection process the agent sends the information to the server application. Prior to this action, the agent executes a simple, yet efficient, data validation that is based on the map of the structure that is sent by the server before the collector agent boot process.

The server application, before being able to receive any position data, must load the floor map that will contain the trackable objects, from a XML file. The maps are modeled as 2,5D worlds and include several structures that can be easily adapted to many types of spaces. Before using it, the server validates the map against a XML Schema.

For instance, a XML file representing a traditional retail shop includes entry areas and an exit one that in this case is designated as a payment area. There are also walls that have infinite height positioned around the map. In the supermarket example these last are named as shelves as can be seen on Figure 2. The proposed XML structure allows specifying a color and a designation for each half of the wall. The half is determined by the largest dimension, and in case of a square it is assumed to be vertically aligned as can be seen on Figure 2.

For each set of received coordinates, the server stores them in a database for universality sake. The
system executes it in a completely completely agnostic way concerning the database provider. At this point it is also relevant to state that the server is a multi-threaded application – having a thread per tag. Each thread writes into a reserved memory location the tag’s current position. The previous position is not overwritten; instead it is stored in the database. Each thread includes some recovery proceedings such as deleting all the records of a specific tag in a given time frame if it stops transmitting its location for a long period of time regarding the context at hand.

As each thread executes, the server GUI is able to display online data. This data represents where the tags are located on the map. Other types of views involve processing the data in real-time, using computer memory and/or by consulting the database registers for further in depth reports (section 3.2).

3.2 Real-Time Tracking Visualization and Concentration Maps

The Server GUI includes several different views of both the online and historical data. In all these views there is a visual representation of the map. In order to draw a map the server requires to systematically perform a scale transformation involving real world coordinates and pixel coordinates. This kind of transformation must be dynamic because, in any instance, the GUI can be resized. When the view involves positioning tags on the map this transformation is also applied to their centre positioning.

The simplest view allows representing online the tags in their actual positions. Other views are obtained through the server’s knowledge extraction features. The Zone Matrix consists in determining in real-time which are the most and least populated zones, denominated as hot and cold zones.

It is also possible to consult this data on a wider time frame considering the same space with or without the same layout. This last feature requires database access. The zones are automatically obtained by dividing the space into a grid with flexible dynamic resolution. This dynamic division allows both a more in depth study of the hot and cold zones and also a less detailed one in order to study, for example, the spaces quadrants occupation.

Several other results are obtainable by accessing memory-based data structures for limited time frame analysis and by querying the database in similar modes. It is possible to obtain historical client paths, the shop areas walls, shelves or objects that were more observed by the clients.

In order to reproduce historical client path recognition all the clients coordinate are stored into the database with a timestamp that is related to a given map. Figure 3 exposes the paths taken by several clients in a given time frame.
of these directions the observer throws a *vision cone* having its center coincident with the direction at hand.

The first intersected walls are considered to be visible, and the others invisible. The Figure 4 summarizes with descriptive colors which were the shelves that were most observed by the moving targets.

![Figure 4: Visible Walls with Historical Data.](image)

4 RESULTS

In this section, the project’s results will be depicted taking into account three generic levels: system’s main features; simulation statistics extraction and global application; and architecture stability and feasibility.

For simulation purposes, there had been considered two standard computer configurations: a high-end machine with 4GB of RAM, an Intel Core 2 Duo E820 CPU and a SATA II 320 GB 16 MB cache hard disc denominated as configuration A, and a low-end configuration, denominated as B, equipped with a 2 GB of RAM, an Intel Pentium D 3.00GHz and a SATA II 250 GB 8MB cache hard disc. Both systems were equipped with Windows Vista Ultimate and the simulations were performed with similar workload conditions. In Figure 5 the experiment’s results are fully depicted as for both hardware configurations four different scenarios were simulated. For each one it was recorded the CPU time needed to perform the most demanding task – real-time dynamic grid concentration levels with memory-based historical data – and the presence of absence of image flicker, with a different number of tracked items ranging from a single one to one thousand.

As previous note, one shall point that for single item tracking, the measured CPU time for both configurations is not available as the benchmarking tool reported zero seconds. The results showed that for ten items, the differences between low-end and high-end computers is absolutely negligible. For one hundred items, configuration B needs twice the time of configuration A but real-time visualization is not jeopardized in anyway. In both cases, for this scale there was not registered any flicker effect and the process time was compatible with a real-time system. Only if the scale is pushed to one thousand, configuration B takes three point seven seconds to compute and even configuration A takes two point three seconds. These figures show that for this kind of scale it is needed a high-end computer system – even if one consider the traditional consumer market products – and hard real-time requirements are not met but one might still assume near real-time features that are perfectly adequate for this kind of management/monitoring systems.

![Configuration A Configuration B](image)

<table>
<thead>
<tr>
<th>Number of Items</th>
<th>Configuration A</th>
<th>Configuration B</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>N/A</td>
<td>N/A</td>
</tr>
<tr>
<td>10</td>
<td>15</td>
<td>23</td>
</tr>
<tr>
<td>100</td>
<td>125</td>
<td>235</td>
</tr>
<tr>
<td>1000</td>
<td>2300</td>
<td>3700</td>
</tr>
</tbody>
</table>

Figure 5: Simulation Performance Benchmarking.

Regarding the first aspect, all the enunciated predicted functionalities, technically described in the previous section, were successfully implemented and fully tested. As illustrated in Figure 6, it is possible to visualise in real-time the location of up to one thousand items overlapped with the shop floor layout. This number of items can be increased but it is dependent of the external location system’s features. In the same illustration, it is visible the tool’s flexibility in what concerns to layout management and design as all shop floor static structures are fully defined and described through a simple, yet flexible XML configuration file. With this approach, it is possible to model heterogeneous environments and, therefore, apply the proposed system to several domains.

In spite of the importance of the mentioned functionalities, the greated added value resides in the knowledge extraction extendend features. Having in consideration the online item location gathering, the system is able of real-time item path reconstruction and visualization operating both in memory-based or database access, depending on data dimension. Conducted experiments showed that real-time memory access is feasible using a low-end computer – with 2 GB of RAM – for tracking one hundred items for a period of an hour at a medium pace.
As illustrated in Figure 7 the location data is used to extract more significant information about item dispersion/concentration both in the present and also considering historical data. It was used a gradient scale where concentration levels vary through the RGB scale where red means high levels of concentration and blue low levels.

One significant functionality of this model is the possibility to perform calculations based on completely flexible and dynamic projection grid. This option proved to be efficient on online data processing for a significant number of tracked items - approximately one hundred - without database access by using a temporal location matrix. This feature enables a full detailed concentration analysis in real time when recurring to a high definition grid that divides the layout in small areas; and enables swift high big-picture studies when using a less tight net. This capability is appropriately described in Figure 8.

Finally, considering the features results description, one shall paint the relevance of the vision module.

This application requirement performs the emulation of an omnidirectional vision of each trackable item. The described algorithm is able of identifying the visible objects, having into consideration both single instantaneous data and historical information, previously collected and stored – in direct memory access or in a database. In the conducted experiments, this system’s module also showed high levels of efficiency and correctness; much similar to the ones already described in the above paragraph.

Regarding the system’s global architecture definition and implementation, the undertaken simulations demonstrated its adaptive capability through its flexibility in what concerns to both the database provider and, perhaps more important, to the location system technology. These characteristics greatly enhance the whole system’s applicability in several scenarios. Still in this domain, the distributed system’s design enables the usage of low-end computers. Therefore it constitutes an incentive to client’s IT infrastructure reusage while minimizing the solution’s economical impact. Simultaneously, this approach enables greater site manager’s empowerment through real-time information access to all system’s features visualization. These actions can be triggered for both partial and global organization providing more and deeper analysis points of view.

5 CONCLUSIONS

Considering the project’s simulation environment and the achieved results, one shall state that, although the location engine had been implemented in order to realistically simulate traceable items, all concept has been demonstrated. The developed prototype proved to be efficient and effective in large scale distributed data gathering and real-time item location visualization.

Taking into account the system’s architecture, it was verified the concern in allowing multi-store management with both distributed modules and central integration concerns that enable consistent and online knowledge extraction and visualization. Having in mind the different application modules, one ought to refer that the integration with the XML-based layout manager proved to be extremely flexible to accom-
moderate distinct real scenarios and yet realizable and realistic.

Considering the most noble project’s slice, one ought refer to the previously enunciated knowledge extraction features. Having as support the results depicted in the last section, it is secure to state that the system is able to produce real-time shop-floor visualization maps with intelligible data on online item localization; individual item complete path routes; online and historical population density rates and path routes concentration; and also item-vision enabled concentration maps – as emulation for item omnidirectional vision, yet considering occlusions. All of these features are allowed for graphical user interface through different grid dimensions for distinct analysis granularity. Bearing in mind the project as whole, one shall state that the developed knowledge extraction platform with online and diversified visualization tools constitutes a solid ground for online item tracking and heterogeneous space management with distributed capabilities. One final major advantage of the proposed system resides in the total independence regarding the external position engine both in terms of suppliers and even more important in terms of base technology.

In spite of the enunciated project’s accomplishment, an even by being in prototype stage – yet reliable and fully functional – there are several future work areas that are able to greatly enhance the system’s appliance and success. From these, the most relevant ones are believed to be centered in eccentric shop-floor layouts both in terms of shape and multi-level buildings; complete path routes analysis enabling common node fusion for global paths probabilistic construction; flexible and dynamic report definition tool with configurable alarm triggering; and perhaps the most interesting would be the characterization of ‘what-if’ scenarios with simulated traffic based on real historical data. Considering both the project current achievements and the depicted future work areas, one might identify the most desirable implementation domains. Although there are not limited to, the proposed system might be useful for traditional retail environment for shopping cart tracking; tracing clients through a commercial area such as shopping centers; enabling item tracking and route analysis in an hospital; producing activity reports and analysis in controlled areas such as penal complexes, mental institutions or closed educational organizations.

ACKNOWLEDGEMENTS

The authors would like to thank Professor Eugnio Oliveira for inspiration and the contribution regarding behavior pattern recognition and artificial intelligence guidelines, Professor Augusto Sousa for enlightenment in the portal cells algorithms.

REFERENCES


A FRAMEWORK FOR DYNAMIC KNOWLEDGE ACQUISITION

Ana Aguilera and Alberto Subero

Facultad de Ciencias y Tecnología, Universidad de Carabobo, Valencia, Venezuela

aaguilef@uc.edu.ve, asubero@uc.edu.ve

Keywords: Collaborative Work, Multi-agents, Medical diagnosis.

Abstract: This article proposes a conceptual model based on agents in medical workgroups for the resolution of clinical cases. Our objective is to provide a suitable conceptual framework for the dynamic acquisition of expert knowledge from different sources. This framework is sufficiently flexible to allow its transformation to the symbolic level that finally characterizes the implementation approach; it is a design model prior to implementation. We considered an interface design to be used for the integration and combination of different sources of expertise. The subject which interests us is the knowledge produced in the workgroup processes.

1 INTRODUCTION

There are many applications related to the workgroup such as decision-making, monitoring, transmission, planning, idea generation, problem resolution, resulting discussion, negotiation, conflict resolution, analysis and system design and collaborative group activities such as the preparation and distribution of documents, to mention just a few. This work style is illustrated by real-life applications, where their nature is clearly collaborative. Several of these applications imply various specialized fields, such as medical diagnosis. The field of medicine has become so vast that it is divided into several specialities and many medical cases are considered by group analysis. Among the various workgroup activities, complex-problem resolution is highlighted as being an increasingly important subject today.

Different information technologies which consider these group modalities have been developed in software engineering. Thus, we found a broader field entitled group support systems or electronic meeting systems which include other wide areas like Group Decision Support Systems (GDSS), Distributed Group Support Systems (DDSS), Computer Support Collaborative Work (CSCW), Groupware, where the common denominator of such technologies covers, but is not limited to, distributed facilities, computer hardware and software, audio and video technology, procedures, methodologies, facilitation, and applicable group data (Turban, 1995).

On the other hand, the evolution of the application domain in traditional Artificial Intelligence (AI) to cover other complex and heterogeneous fields such as aid to decision-making, form recognition and comprehension, process control, etc., shows the limits of traditional AI approaches. All this has contributed to the birth of a new discipline: Distributed Artificial Intelligence (DAI), which is interested in the intelligent behaviours (and their modelling) that are the outcome of cooperative activity between several agents. Nevertheless, the passage from individual behaviour to collective behaviour is not only considered as an extension but as an enrichment of AI, as new properties and new behaviours emerge from it. The purpose of DAI is to provide a remedy for the shortcomings of the traditional AI approach by proposing expertise distribution for an agent group capable of working and acting in a common environment, and solving possible conflicts. New concepts in AI have appeared, such as cooperation, action coordination, negotiation and emergence. Three fundamental research areas of DAI are multi-agent systems (MAS) (Weiss, 2000), distributed problem solving (DPS) and Parallel Artificial Intelligence (PAI).

2 THE PROBLEM

Everyday medical work (implying inter-consultation between specialists, case conferences, and hospital morning rounds) includes exchange and cognitive
processes within groups. It is possible to observe different physicians (specialists and non-specialists) working together, at the same time or at the different time, on the same patient (clinical examination often requires the cooperation of several specialists). The contribution of these exchanges to solve complex problems is becoming more and more essential. These acts of collaboration are important for clinical decision-making concerning diagnosis and treatment both for students still in training and recently qualified physicians, without forgetting the continuous experience which helps to develop expert knowledge. There are several examples: cases of multi-system illnesses where physiopathology and the nature of origin of the illness make it necessary to examine the results of several diagnostic procedures; patients with chronic disorders such as diabetes mellitus, obstructive pulmonary illnesses, cardiological illnesses; or patients receiving palliative care at home.

On the whole, the most frequent scenarios where it is possible to observe this collaborative pattern are:

Inter-consultations. This is a process where a physician needs specialized consultation with several other physicians. There are two possibilities for this process: 1) between specialists with the same speciality (consultation between radiologists who study an image to decide on a diagnosis) and 2) between specialists with different specialties (an obstetrician who refers his/her patient to a cardiologist for a coronary problem).

Case Conferences and morning rounds. There are cases where it is possible to find several physicians (specialist physicians interconnected with general practitioners) exchanging information on several medical cases in order to make patient evaluations, to work on the publication of cases, etc. These are considered as staff meetings.

A clinical case illustrating this type of work is shown in figure 1 (A more detailed explication is found in (Quintero, 2003). In this case, we can observe the interaction between a generalist, a radiologist and a pathologist.

3 PROPOSITION

3.1 Architecture

The problem follows a structure on three levels (Figure 2): the collaborator level or users (human agents), computer systems agents (software agents) and the data level (databases and knowledge bases). This architecture determines the structure on which the framework will be defined and permits us to clarify the semantic definition of the problem under study.

3.2 Methodology

Different methodologies for the development of knowledge-based systems (where knowledge acquisition and modelling play a leading role) have emerged during the past 15 years. We can mention in particular methodologies such as KADS (Schreiber, 1993), VITAL (Domingue, 1993), REFLECT (Reinders, 1991), (van Harmelen, 1992), ACKnowledge (van Heijst, 1992), CommonKADS (Schreiber, 2000) and KATEMES-MultiExp (Ladibi, 1995). KADS and VITAL focus on the development cycle of a knowledge-based system, REFLECT proposes theories and architectures for reflexive systems, ACKnowledge introduces directive models for knowledge modelling. CommonKADS offers methodology for knowledge-intensive system development and for knowledge management and KATEMES-MultiExp is a prototype tool for multi-expert knowledge acquisition.

The approach used is based on an extension of the KATEMES-MultiExp methodology. This method integrates a set of models for modelling multi-expertise and its objective is to help the knowledge engineer in the collecting and modelling phase of human behaviours in cooperative problem solving. These models are: agent, organization, cooperation, task and communication models (Aguilera, 2003b). We added a coordination model (Aguilera, 2003a) because originally KATEMES was designed for knowledge acquisition before system development and we considered the dynamic knowledge acquisition approach. One of the important features considered is the ability to extract expertise in a non-intrusive way. This minimizes biased behaviour by monitoring expert decision makers directly (through the information system) during their daily tasks, without changing anything about their choices or strategies. The approach is based on the concept of the agent, with the idea of modelling a group of experts via a community of agents in interaction. This gives us the additional advantage of multi-agent systems which provide a way to relax the constraints of centralized, planned, sequential control and to make systems that are decentralized, emergent and concurrent available. We are convinced that, the modularity of a multi-agent architecture facilitates knowledge acquisition and the parallel design of each expertise. In fact, the multi-agent systems are
Circumscribed 2-3 cm Nodule located in the right upper lobe of the lung with the presence of interior calcifications of non-specified type. Scanner recommended. Heart and rest of study without obvious modifications.

**CAT Lung scanner:** It indicates a 2 x 3 cm mass with non-epimacular aspect located in the right upper lobe of the lung with non-specified calcifications. There is no affection of Mediastinum lymphatic ganglia. There are no other masses in the thorax.

Pulmonary biopsy: macro and microscopic analysis of post-operative piece.

**Diagnosis:** ENDODBRONCHIAL HAMARTHOMA.

The patient leaves the hospital and considering the benign origin of the pathology, the doctor recommends an annual check up with his general practitioner.
medical appointments, takes part in his/her cure, being takes part in his/her treatment. He/she requests patient).

observe only one part of the “outside” (i.e., the about the patient. Each one of these specialists can divide the tasks amongst themselves and share data knowledge of agents is heterogeneous. The agents neurologists, pathologists and others. The specialities, i.e., human agents such as radiologists, classification based on the different medical the case).

The communication protocol between agents follows conversations; that is, somebody intervenes and the human communication procedure based on personal mechanisms.

The framework definition is given, simultaneously in the structuring and conceptualizing of each model in the extended KATEMES-multiexpert methodology. At the conceptual level we can define the following models:

- Agent model
This model enables us to identify and define all human and artificial individuals interacting and taking part in cooperative problem solving. These agents are (figure 2):

- The medical meeting agent: This is considered as a high level hierarchical agent. It is a set of agents and it is defined to identify the group of physicians. The communication protocol between agents follows human communication procedure based on personal conversations; that is, somebody intervenes and the others remain silent and listen, and when one person has finished, another intervenes and so on. The leadership position can be assumed by the doctor sitting opposite the patient (Physician in charge of the case).

- Medical speciality agents. They define a classification based on the different medical specialities, i.e., human agents such as radiologists, neurologists, pathologists and others. The knowledge of agents is heterogeneous. The agents divide the tasks amongst themselves and share data about the patient. Each one of these specialists can observe only one part of the “outside” (i.e., the patient).

- The patient: A human agent who as an active being takes part in his/her treatment. He/she requests medical appointments, takes part in his/her cure, asks for information. He/she makes it possible to establish the clinical context.

- The planning agent: An artificial agent which manages everything related to time in the system: assigns the tasks with a time context, checks the time of their activation, execution and ending. It returns the requests of task assignments if it has not received an opportune response. It centralizes information about task executions and their different states of development.

- The distributor agent: An artificial agent which locates the agents that will execute the tasks, according to their specialities and availability for executing them. For this, it takes into account the urgency of tasks required. It centralizes information about group members.

- The manager agent: An artificial agent which controls the requests for tasks. It sends and receives results.

- Interface agent: Artificial agent who controls the security accesses to systems and data. It adequately adapts the appropriate interface according to each user with his/her different levels. It centralizes general access information.

- The data agent: An artificial agent which manages data in the databases and knowledge bases. It manages metadata, user view definitions and authorizations to data access. It also controls anything related to data integrity and recovery mechanisms.

Subsequently and for reasons of brevity we will define only the radiologist agent. This is a human agent, an expert and a specialist belonging to the radiology department. He may be or not be subordinate according to his status. He is part of a community including radiologists or other colleagues of different specialities (general practitioners, pathologists, etc). All of these people rank equally and take part in the development of a diagnostic solution. This agent also has a hierarchical relationship with the department head. With respect to inter-agent relationships, the radiologist could influence the reasoning line of an agent with another speciality, who requires his service for a diagnostic examination, or he could directly influence another radiologist who discusses a particular case with him. In the radiology department, software systems for image processing are frequently used. The execution of these systems can be initiated or stopped at any time by radiologists. Concerning cooperative relationships, the agent can assist another radiologist in the establishment of a diagnostic hypothesis or in other tasks. In the medical group context, his task does not interfere with other tasks but sometimes its
accomplishment is important for the continuation of other tasks. Other inter-agent relationships can be established, but in order to simplicity, only those related to the example given (figure 1) will be mentioned.

- **Expertise model**
  Knowledge based systems are founded on a separation between the knowledge necessary for problem solving (domain knowledge) and the reasoning mechanisms needed to exploit this knowledge (inferential process knowledge for problem solving). In the research field, this subject remains open and there is a great diversity of analyses and different viewpoints about it. This topic is related to reasoning mechanisms used by doctors in diagnostic problem solving. Many variables are considered: patient, doctor experience, general medical knowledge. The discussion of this theme is extended and goes beyond the domain of this article.
  We suggest readers refer to specialised literature (Torasso 2001, Long, 2001), in particular, we refer to (Park, 2003).

- **Organisation model**
  In the architecture proposed (Figure 2), we find artificial and human agents. From this, we identify the medical group as a hierarchical structure with two levels; headed by the physician treating the patient, and then the rest of the doctors are considered to be collaborators. This structure can change by establishing as group leader another doctor in any other speciality.
  It is also possible to observe other hierarchical structures, for example in a medical group within the same department (radiology department, pathology department, etc.), with the department head at the highest level.
  Within a group of artificial agents, some have only coordination functions. A hierarchical structure is present and centralized in the manager agent. It is possible to observe delegation of tasks. For example, the manager agent requests that the planning agent specify the localization of a task and then the latter asks to the distributor agent which finally assigns the task.

- **Cooperation model**
  The model presupposes the cooperative motivation of its participants. In the case of a medical group, the doctors can be assisted within the same speciality, when they request consultation at any step of the diagnostic process (doubts, confirmations, etc.). In different specialities, no one interferes directly with the others but their effective participation, in the total establishment of diagnosis, is decisive.

- **Task model**
  To arrive at a diagnosis, doctors carry out several tasks. In particular we mention: hypothesis generation, selection and test of hypotheses and diagnostic conclusion. In the case of hypothesis generation, which is the example presented here, it is not clear if the same process is followed equally in every medical speciality. Undoubtedly, there are different contexts and expertise models can vary within the same speciality. However, it is not our objective to present a discussion of the diagnostic process.

- **Communication model**
  There are several communication mechanisms between agents given by the type of their interrelations. For example, in the communication between human agents, there are communications based on spoken conversations, written conversations (chats) or by data (medical file). Between human and artificial agents, the communication is established by an interface agent. Among artificial agents, there are mechanisms based on message passing, and finally, between artificial agents and data, through a data agent. Communication modelling is described by inter-agent relationships via R-intervention relationships which do not contradict the organisational structures described. Some of them are shown in the definition of the radiologist agent.

- **Coordination model**
  Coordination is supported by artificial agents: the manager, the planning and the distributor agent. They centralize everything related to the coordination tasks required. Coordination will also depend on the dynamics of the cases exposed, their characteristics, their emergency requirements, their resources and the interrelations established between the active agents.
  Therefore services offered by this model depend on: the session context, the resources provided by participants and the social behaviour of participants (actions carried out by human agents) (Ossowski, 1999). The combination of these three elements creates the dynamics of coordination and the basic actions for this collaboration; the context fixes the limits of these dynamics.

### 3.4 System Design and Implementation

At implementation level, we design a multi-agent architecture based on the framework proposed. The different medical specialities are modelled well with a modular structure. The current project is progressing. So far, we have designed and implemented a groupware application that facilitates the management of clinical cases (Cárdenas, 2001) and a groupware application for radiological
teleconsultations. The conceptual data model of our application uses an entity relationship model. At the moment, we are developing a workflow application for the coordination of medical diagnostic work.

4 CONCLUSIONS

In this article we defined an agent-based conceptual framework that integrates the social aspects of human activities, the cognitive aspects structured in databases and knowledge bases of the system, as well as, the computer system aspects which are present in the software applications and information technologies, in the medical diagnostic context. This approach is model-based; thus six models are defined: the agent model, the organization model, the cooperation model, the task model, the communication model and the coordination model. It is possible to instantiate these models for particular situations, for example, clinic and hospital frameworks. We have considered the extension of the metamodel proposed in (Ladibi, 1995), the annexing of the coordination model and we provided general templates for medical groupwork. The selection of the model-based method and the design of the model were directed towards the dynamic knowledge acquisition produced in collaborative medical interactions, which is the next step in this research.

ACKNOWLEDGEMENTS

We are grateful to FONACIT, Venezuela, under Project No G-2005000278, which has sponsored this work.

REFERENCES

S. Ossowski, “Coordination in Artificial Agent Societies. Social Structure and Its Implications for Autonomous Problem-Solving Agents”, Number 1202 in LNAI. Springer Verlag, 1999.
Keywords:  Reputation systems, evaluation, resource, Web community.

Abstract: The purpose of the PRES model is to build a flexible and easy way to manage resources in a personalized manner. Our proposed model assures for every user that his preferences are important and permits the formation of some homogenous groups on the basis of these preferences. The homogeneity is due by the relations resulted from the explicit and implicit evaluations of resources. The purpose of the proposed model is to build a flexible way to filter irrelevant resources for users. In this way, a user which is member to a community based on the PRES model will dynamically see information that he/she is most interested in.

1 PREAMBLE

In this moment the WWW space stores large amounts of data which are continuously growing. The main problem that appears is to find solution to use efficiently the existent resources.

A first step to solve this problem is to associate metadata to resources. As a fact, it is a manual classification process performed by the user (E.g. delicious, digg.com). This direction is a part of explicit Web that is realized through explicit activities as tagging or digging.

An important direction, using the above solution, is to obtain data/information by observing and analyzing the user actions. Thus, we enter the space that is known as implicit Web (O’Reilly, 2005). An important drift of it is collective intelligence domain (T. Segaran, 2007).

In this context, this work proposes the analysis and projection of a prototype of a reputation’s personalized evaluation system in a Web community (PRES - Personalized Resource Evaluation System). The originality of this approach consists in the chosen perspective to accomplish the evaluation.

This work is structured as follows: in section 2 we describe a short survey on the present situation (O’Reilly, 2005, H.Zhuge, 2008). In the next section we present the problem, we explain why such a system is necessary and we present the proposed model. In the fourth section we analyze the benefits of the proposed system. The article ends with an overview on the discussed domain, mentioning the future directions.

2 ACTUAL SITUATION

At this moment there are many sites that collect various information about thousands or even millions of people on the Web. This information is obtained often without even interrupt user actions with questions. His behavior and profile can be obtained from this information using different techniques like machine learning and statistical methods.

In the collective intelligence spectrum we have two different approaches, one exists due the information furnished by users (e.g. Wikipedia). The other part of the spectrum is based on different algorithms which allow obtaining new information that enhance the user experience. An important example in this sense is Google, which uses links to rank web pages, but also collects and process data obtained from situations when advertisements are clicked.

Other examples consist of web communities that use recommendation systems (Massa, B.Bhattacharjee, 2004). In this cases there are
collected information like purchasing history and user characteristics, and the system make proper recommendations based on them (e.g. Amazon, Netflix).

Other examples consist of web systems which use reputation systems (Golbeck, Hendler, 2005). Reputation systems are extremely useful in those communities where the users have to interact with some resources posted by other users or they have to interact with other users. (E.g. YouTube, Slashdot, Flicker). In these situations, using experience of other users would be very useful. Also, reputation systems are useful in setting some evaluation levels for users and resources (e.g. more or less interesting resources). There are a variety of reputation systems. A well-known system, mentioned before, is Google Page Rank (A. Langville, C. Meyer, 2006) that is based on complex algorithms that assure the web page ranking.

Another reputation system is that used by eBay. The system assures a feedback profile for each member.

Each feedback consists of a positive, negative or neutral value (these values are obtained from the ratings of the transaction partners) and a short comment.

Everything2 is a knowledge base that contains reputations system both for users and their posted articles. The system is based on anonymous votes of other users which determine positive or negative ratings. Negative evaluated articles are deleted. The users are evaluated on the basis of the number of their submitted articles (and not deleted) and on the average of their associated values.

Such a system implies some problems: new users posting articles that receive negative feedbacks may appear. These articles will be deleted, thus discouraging new postings by such users. Even the experienced users hesitate to post new articles which they consider as being not very good, because the received negative feedbacks are not deleted. Also, in this kind of system the re-actualization of older articles is less appreciated.

Slashdot has a reputation system named karma. In this system there are moderators that can make the evaluations in a similar way to the system Everything2. Every user may become moderator if he has a good karma obtained on the basis of the ratings associated to their comments. But this moderator state is temporary until he uses the available votes. This evaluation system is criticized because it is weak on issues like Anonymous Coward or sock puppets (R. Falcone, S.Barber, L. Korba, M. Singh, 2002).

Another system we referred here before is Wikipedia that represents an online community containing a great number of users, but not using a formal reputation computation mechanism.

As in the previously discussed systems, a less visible user hierarchy exists. All users, on the basis of their contribution, may receive the so-called barnstar acknowledgement. Although one can follow each user posting history, it does not exist a particular rating system.

3 PRES MODEL PROPOSAL

3.1 Context

In section 2 we have discussed a set of reputation systems (R. Falcone, S.Barber, L. Korba, M. Singh, 2002), but in all these related approaches we do not find a personalized evaluation. In this section we explain what a personalized evaluation means, from our point of view.

In a Web community there exist a lot of resources. There are human resources and other types of resources. The people have either different or similar profiles. Therefore, they are interested in either different or similar resources.

We quantize this interest with values which are provided by the user for other users or resources. Also, this interest will have an indirectly computed component. We give a simple example here, the other cases being analyzed in section 3.2. We have the situation when a user evaluates favorably one or more users. These users evaluate favorably a given resource. Even if the user does not evaluate directly that resource we will consider an implicit favorable evaluation (J.Golbeck, J. Hendler, 2006). Thus, the user has the chance to access more relevant resources for him.

In our system there does not exist an absolute value of good or bad resource characteristic. A resource can be good for a set of users but not useful for other set of users.

In section 3.2 we establish a set of metrics (J. L. Mui, 2002), taken into account by the evaluation mechanism, for the purpose of measuring the usefulness of a resource for a given user.

Whenever new users become community members they can interact with the users corresponding to their preferences. Also, they will be able to access much faster the proper resource set. This represents the general direction our system is based on.


3.2 The Proposed Model

First we define the vocabulary used in the developing model. We also specify the used notations and their semantics. The system will contain:

- Users which know other users.
- The list of the users considered to be interesting for a user.
- Users nominated by a community as evaluators. We use notations E₁...Eₙ to indicate the community evaluators. These evaluators are in fact some reviewers. They will be useful for the new users which have not established their own knowledge list yet.
- Known person list of a user. Initially, it contains the community reviewers list only.
- Resources – their definition is made accordingly to the definition given by (T. Berners-Lee, 1998).

So, in our system one considers as resources everything having an identity (e.g. electronic document, an image, a service and eventually a collection of other resources). There are considered as resources those that cannot be accessed directly via Internet (e.g. human beings, organizations)

- Worth – this parameter is a metric. This metric signifies a given rating, according by a user to a resource or a user. Also, the worth can be obtained (quantized) indirectly.

This parameter – Worth – takes the following values:

1 = useless/spam
2 = poor
3 = worth attention
4 = good
5 = exceptional. We note this limit with MaxWorth.

We think of using the 1-10 interval for possible values for Worth metric, this approach assuring higher granularity in resource evaluation. We prefer the above specified selection to simplify the model. In future works we will analyze if this aspect has a major influence on the resource evaluation manner.

We will use a set of constructions which have the following associated semantics. In fact, these constructions can be mathematically considered as functions (eventually partial functions) or, from the implementation point of view, they are considered associative tables:

- Explicit worth of a resource: \( WE_{UR} (user, resource) \) – explicit worth, represents the rating for a resource, this rating being given manual by a user
- Implicit worth of a resource: \( WI_{UR} (user, resource) \) – implicit worth, represents a rating inferred from the set of existing ratings from the known person list of a user
- Explicit worth of a user: \( WE_{UU} (user, user) \) – explicit worth, represents the rating for a user, and the rating is given manual by the user to another user
- Implicit (deducted) worth of a user: \( WI_{UU} (user, user) \) – measure how close are his preferences to the others preferences (The preference can be considered: the accepting degree of a point of view or the appreciation degree of a piece of art).

Implicit we consider that we have:

\( WI_{UU} (user, evaluator) = MaxWorth; \)

If an user evaluates an evaluator in an explicit manner, then this evaluation - \( WE_{UU} (user, evaluator) \) – will have priority.

- we consider the function \( WU(user, user) \) for every pair of \((user, user)\) from a Web community

Its value will be \( WE_{UU} (user, user) \) if there is an explicit evaluation (different from 0), otherwise its value will be \( WI_{UU} (user, user) \). So, let us consider the users: \( U_x \) and \( U_y \).

If the user \( U_x \) evaluates explicitly the user \( U_y \) then the function \( WE_{UU} \) has a value different of 0 and the value of \( WU(U_x, U_y) \) will be \( WE_{UU}(U_x, U_y) \). If \( U_x \) does not make an explicit evaluation for user \( U_y \) then \( WU(U_x, U_y) \) value will be the inferred value which is actually the value of \( WI_{UU}(U_x, U_y) \).

We consider a function \( WR(user, resource) \) for every pair \((user, resource)\)

\( WR (user, resource) \) value will be \( WE_{UR} \) if the user evaluates explicitly the resource, thus the value of \( WE_{UR} \) exists. Otherwise \( WR \) value will be the value of \( WI_{UR} \).

Therefore, let us consider the user \( U_i \) and the resource \( R_j \). \( WR (U_i, R_j) \) value will be \( WE_{UR} (U_i, R_j) \) if user \( U_i \) has explicitly evaluated the resource \( R_j \). Otherwise \( WR \) value will be \( WI_{UR} (U_i, R_j) \) if the user \( U_i \) did not evaluate the resource. This value is based on the ratings to \( R_j \) made by users which are in known list of the user \( U_i \).

We will define the manner of computation of the implicit values introduced above.
Implicit WI_UU Value Computation

Let us consider two users \( U_i, U_j \) from the Web community. In order to define \( WI_UU(U_i,U_j) \) we introduce the following partial functions:

- \( WIU_UU(U_i,U_j) \) – its value indicates the deducted worth based on explicit evaluations made by users to each other
- \( WIR_UU(U_i,U_j) \) – its value signifies the deducted worth based on evaluations that users do to the same resources

Defining \( WIU_UU \) value on the basis of the explicit values

Let the users, whom we have explicit ratings from user \( User_j \) to be \( U_i, 1 \leq i \leq k \), be \( U_i,...,U_k \).

Therefore we have the definition \( WE_UU \) (\( User_i, U_j \)). Also we have explicit ratings from \( U_i \) to \( User_j \), so we have defined \( WE_UU \) (\( U_i, User_j \)) (see Figure 1).

![Figure 1: WI_UU computation based on explicit evaluations.](image)

The need of partial function \( WIR_UU \) when the set of users used for defining of \( WIU_UU \) is the empty set. This means that we do not have a set of users \( U_i, ..., U_k \) whom we have explicit ratings from \( User_j \) to \( U_i \), \( 1 \leq i \leq k \) and also we do not have explicit ratings from \( U_i \) to \( User_j \). In this case we can obtain information on the basis of the worth of a set of resources evaluated by users. These resources are required to be evaluated by both users. Thus, on the basis of the evaluations of the same resource, one can obtain a mutual evaluation of two users.

Let us consider: \( U_i, U_j \) and the resources \( R_1,...,R_n \). If there exists \( WE_UU(U_i,R_j) \) and \( WE_UU(U_j,R_i) \), for all \( 1 \leq i \leq n \), then the value of \( WIR_UU(U_i,U_j) \) will exist and it will be equal with \( WIU_UU(U_i,U_j) \). We define \( WIR_UU \) (\( U_i,U_j \)) as follows:

\[
WIR_UU(U_i,U_j) = \text{MaxWorth} - \frac{\sum_{i=1}^{n} |WE_UU(U_i,R_j) - WE_UU(U_j,R_i)|}{n}
\]

where \( 1 \leq i \leq n \). The demonstration of the assertion:

\[
WIR_UU(U_i,U_j) = WIU_UU(U_i,U_j)
\]

is obvious. Therefore, in the case when we want to obtain \( WIU_UU \) on the basis of resource evaluation \( WIU_UU \) has the value of \( WIR_UU(U_i,U_j) \).

Finally the worth of \( WIU_UU \) (\( U_i, U_j \)) will be \( WIU_UU(U_i,U_j) \), if defined, or \( WIR_UU \) if defined, or it will be a default value fixed in the system configuration.

In a future work, we will present a mechanism to obtain a complete function \( WIU_UU \) without this implicit value. In addition, if we have both user-user and user-resource evaluations, then we can foresee a given priority between them.

Implicit WIR_UU Value Computation

We will define the manner to obtain the worth of the implicit evaluation \( WIR_UU(U_i,R_j) \) – whom a user \( U_i \) provides to a resource \( R_j \), \( 1 \leq j \leq n \), where \( n \) is the resource number of the system. We consider that the user \( U_i \) has in his known person list the following users: \( U_{i1},...,U_{ik} \). These users have evaluated the considered resource. This means we have defined the following relations: \( WE_UU(U_{ij},U_j) \), \( 1 \leq i \leq k \) and \( WE_UU(U_i,R_j) \), \( 1 \leq i \leq k \).

The implicit rating provide by \( U_i \) to resource \( R_j \) is represented by the proportion between: sum of the product of the rating weights of the user \( U_i \) for each user from his list and the value provided by him to
resource $R_x$, and the number of users (which is $k$ in our situation)

$$W_{I\_UR}(U_x, R_x) = \frac{1}{k} \sum_{i=1}^{k} PE(U_x, U_i) \cdot WE_{\_UR}(U_i, R_x)$$ (5)

where $1 \leq i \leq k$. We introduce the worth average provided to a resource and we denote it with $WA$. The value of $WA$ (Resource) represents relevant statistical average provided to a resource by all users. $WA$ for a resource inside a Web community plays the same role that page rank plays in Web page evaluation. This metric is necessary in case we do not have enough trustworthy evaluators in the community.

### 4 ASPECTS REGARDING THE PRES BENEFITS

In this section we discuss shortly a set of consequences, due to the way the system has been modeled. We will argue our assertions through few examples and in a next paper we will give the appropriate algorithms used for these cases.

- The system assures the property to see the things prioritized the same way as similar users.
- The spammers will see more spam because the system groups the users by their preferences.

Let us consider a web community with users $U_1, \ldots, U_k$. We can consider that a new user $U_x$ joins the community and posts a new resource - $R_x$.

The resource posted by $U_x$ will be evaluated by the users from community with worth values (implicitly $WE_{\_UR}(U_x, R_x) = 5$).

If $R_x$ is a spam resource, it will be explicitly evaluated by users $U_i$ which are not interested in spam resources with $WE_{\_UR}(U_i, R_x) = 1$ or it will be explicitly evaluated by users $U_j$ which are interested by this kind of resources with $WE_{\_UR}(U_j, R_x) = 5$, where $1 \leq i \leq k, 1 \leq j \leq k, i \neq j$.

Also, let us consider the case when a user $U_i$ evaluates the users $U_j$. Because users $U_i$ have evaluated resource $Rx$ with low worth than the sum $\sum_{i=1}^{k} PE(U_i, U_j) \cdot WE_{\_UR}(U_i, R_x)$ has a low value.

Then the value of $W_{I\_UR}(U_i, R_x)$ will be low and in this case the spam resource $R_x$ will not be considered interesting for the user $U_i$. In other case when user $U_i$ will evaluate users $U_j$ with worth metric with a higher value than the sum $\sum_{i=1}^{k} PE(U_i, U_j) \cdot WE_{\_UR}(U_i, R_x)$ has a higher value and in this case the resource $R_x$ will be automatically consider useful for user $U_i$. We argue with this example one case from a set of possible use-cases. We will discuss in detail these cases and the used algorithms in our next paper.

- The resources which are relevant for the user are on top of the list of visible resources. In this moment we know that Google uses Page Rank system. The new resources, even valuable, will reach hardly on the top, because it takes long until they receive links. And worse it is the fact that if they are not on the top, they do not receive links.

Therefore there exists a very high probability that a good resource is not used.

In our system the new valuable resources appear quickly on top when they are evaluated first by an honest community member (one who tries always to evaluate correctly). If somebody over evaluate his own resource and the others rate it with low marks, the mark $W_{I\_UU}$ will drop, therefore those who add resources are required to give right marks.

- The users will be required to do a fair evaluation.

It will not happen like in the eBay system. In this system, one assures a feedback for each user. The feedback value is obtaining from other users evaluations. One observed that the users are afraid of obtaining a negative feedback. For this reason they post positive feedbacks in a high proportion, hoping that they will obtain positive feedbacks.

- The system can be easily integrated in different Web communities.

Let us consider a real community like LinkedIn. There exist in this moment some posted announcements which offer jobs for this community only. Our system would give the possibility that this announcements to be visualized only by the users with a given profile, the announcement being not useful for other users types.

Thus our system makes it more efficient the information management that is visible to the user.
5 CONCLUSIONS

Reputation system gives people information about past activities of the other users. It can enhance an on-line interaction environment by: helping people decide who to interact, encouraging people to be more honest, discouraging those who are not responsible from participating. The actions, the behavior, the user preferences can be regarded as resources on which one can initiate interpretation and processing mechanisms. PRES system allows each user to have its own evaluation of the resources and of the other users. The proposed metrics can be used for implementation in real Web communities.

In this work we have presented the basic elements of PRES model. For the future development of the prototype we will perform a detailed analyzes of the system properties. In a real system the resources are changing in time. This problem will be studied in our system thus foreseeing the possibility that the users can see and change the given ratings.

Another problem related to the reputation computation that will be studied is a complexity of the algorithm of performing the entire calculus in the system. The computation of WR and WU can be easily performed for a proper number of resources and users. For hundreds thousands of users and resources we need a parallel algorithm to compute periodically the WR and WU values.

REFERENCES


L. Mui, PhD thesis: Computational models of trust and reputation: agents, evolutionary, games and social networks, MIT, 2002
Jennifer Golbeck, James Hendler: Reputation network analysis for email filtering, for the 1st Conf. on Email and Anti-Spam, 2004
T. Berners-Lee et al. (eds.), Uniform Resource Identifiers (URI): General Syntax, RFC 2396, IETF, 1998
LinkedIn. http://www.linkedin.com
POSTERS
A STUDY OF THE EFFECTIVENESS OF “WAKE UP ON LAN” AS A MEANS OF POWER MANAGEMENT

Colin Pattinson and Linton Robinson
Innovation North Faculty of Information and Technology, Leeds Metropolitan University, Leeds U.K.
{c.pattinson, l.a.robinson} @leedsmet.ac.uk

Keywords: Power management, wake on LAN, Energy savings in corporate IT.

Abstract: The growing awareness of the fragile nature of our environment, and of the damage that humankind is causing to that environment, makes it necessary to review the environmental impact of all aspects of human activity. One such area is the operation of corporate information technology (IT) systems. The growth in the number and complexity of such systems over recent decades has led to a consequent increase in their power requirements, to the point where for organisations of any size, “corporate IT” will be a major contributor to the organisation’s overall energy consumption. A drive for overall reductions in consumption will, in the words of the chair of the newly formed UK Environmental IT Leadership Team, mean “the IT department becomes the focus of carbon reduction policies”. It is therefore timely to consider the present situation in respect of power consumption within corporate IT systems, and to explore the potential avenues for reducing that consumption. In this paper, we will briefly overview the area of power usage in IT, before reporting on the outcomes of a specific project in which we explored the operation of the “wake on LAN” method in a real situation.

1 INTRODUCTION

The rapid growth in computer use is placing a strain on the UK’s ageing electrical infrastructure. The growth of networked systems within organisations is a major factor, with an increase in the number of servers in a typical organisation from around 4 in the late 1990s to 10 to 15 times that number (Watson, 2006). In October 2006, there were 2,507,278 registered companies in the UK (Companies House, 2006). While we do not know exactly how many of these companies operated with the 60 servers in Watson’s analysis or even as many computers, it is not unreasonable to suggest that a good proportion will have PCs, Servers and other network devices.

Data for energy consumption of IT systems vary according to the exact definition of what is included, and there is variation in reporting methods. Kawamoto et. al. (2000) found “devices connected to the Internet” for “commercial use” were responsible for 2% of the total U.S. power consumption. In Germany over 7% of domestic energy consumption is “computer-related” (Loerincik 2006).

One attempt to save energy has been to limit the power taken by devices when not in use by placing them in a “dormant” state, like the “stand by” mode of many consumer electronics devices. Note: in a further parallel with consumer electronics, “dormant” is not “off”, and some power is still required. The generic term for this activity is power management (PM). PM can bring about energy savings “…of the 74 TWh/yr of electricity (approximately $6 billion per year) … consumed by the Internet in the USA alone … 32% could be saved with effective PM on desktops” (Christensen et. al. 2007), and the equivalent of 250M litres of gasoline per day could be saved if 1M PC users employed PM (Long 2006).

Many current energy reduction activities (including the one reported here) aim at the desktop (user PC) element of IT systems, reasonably enough in view of the numbers of such devices. However, server and network device (switches, hubs and routers) must not be overlooked. Allowing such devices to enter a dormant state is problematic, as typical network performance and security monitoring techniques expect networked devices to respond to a regular sequence of polling packets; indeed some network management tools generate an alarm if a device does not react to a probe request; therefore there is a contradiction. Furthermore,
“[t]he immense increase in networking … is an important factor [limiting the use of power saving] because network activity can keep computers from entering low-power modes” Webber et. al. (2001).

2 APPROACHES TO PM

In the opening parts of this paper, we argued that there is a contradiction between the conventional way networks operate, PM and cost savings. Let us therefore take an overview of the different potential strategies in this regard. We start by giving a brief explanation of each of these techniques (Table 1).

To some extent, this table is rather simplistic; e.g. overlaps between some techniques could combine to effectively provide a PM solution. In this study, WOL has been chosen primarily because of its ease of use and simplicity in implementation. Also, a previous research project successfully used WOL in both wired and wireless scenarios, so we have confidence in the underlying method.

3 QUESTIONS TO ADDRESS

The objective of this research was to determine how WOL operates; hence we begin by exploring how WOL will be delivered in a typical networked scenario. Crucial to the deployment of WOL across a network are questions of the nature and operation of the communication process; for instance, are WOL packets broadcast, unicast or multicast? What are the issues with regard to a ‘connectionless or connection-oriented approach?’ and finally, what are the issues in terms of efficiency and security.

4 THE MAGIC PACKET (MP)

‘Magic Packet’ (MP) and ‘Wake up On LAN’ (WOL), are the same thing. The Magic Packet is broadcast via port 7 or 9 which can be sent using a variety of connectionless protocols, with UDP the most common. (CapaInstaller, 2006). The Magic Packet structure is similar to an Ethernet packet, but the destination Medium Access Control (MAC) address is repeated 16 times within the packet.

WOL encapsulates such Magic Packets inside a broadcast UDP packet. This gives benefits over raw Ethernet frames including Operating System (OS) interoperability.

The complete UDP packet, sent over an Ethernet interface, looks like this: (Spurgeon 2006)

[Ethernet header][IP header][UDP header][Magic sequence][CRCs]

5 TO BROADCAST, UNICAST OR MULTICAST?

Unicast packets are sent from host to host, a one to one relationship. Broadcast packets are one to many, one host communicating with all other hosts. Multicast is a restricted one to many relationship; a
single device can communicate with a specific set of hosts. Our previous work with the MP showed that it can be transferred by broadcast, unicast and multicast, however for the purposes of this paper, we consider a ‘broadcast’ approach, minimising effort (just one message needs to be created), while maximising coverage. By design, most modern routers will block IP broadcast traffic and restrict it to the local subnet, but this can be changed by the network administrator.

6 WOL QUESTIONS

6.1 Efficiency

We conducted a series of tests to send the MP over a wired network with unicast and multicast instructions to wake machines. In these tests, the packet successfully woke up and shut down workstation(s). When this was carried out using a wireless link, we found that in 2 cases, the instruction to sleep succeeded, but the instruction to wake failed if the target machine was in another subnet. We believe that a correctly configured wireless AP in that foreign subnet will rectify this and conclude from these results that the MP is able to produce satisfactory results and to do so reliably.

6.2 Reliability

It is well known that UDP is not reliable (Bhatti 2007); therefore the immediate question might be why use it? Why not invoke a connection oriented protocol with reliable delivery? Once a command is issued to sleep, there is currently no feedback mechanism to indicate that command was received and obeyed. Instead, the network administrator may need to utilise other tools creating extra work and traffic, reducing the potential benefits of PM. Therefore one challenge for developers is to bridge the gap between the command and the response.

6.3 Security

There are a number of security issues within WOL and its implementation (Robinson, 2007). For instance, applying the ‘layered approach’ of most network designs as in the following model:

\[
\text{Total Vulnerabilities} = (\text{PM} + \text{current security}) + (\text{LAN security}) + (\text{WAN security}) + \text{internet security}
\]

means that vulnerabilities are compounded. We also found that “more advanced systems” do not necessarily mean “more advanced PM” a further indication of the need for research and development into PM. Min and Chandrasakan (2003) argue that computational algorithms and low power digital hardware trade energy for quality, because digital processing occurs more slowly and uses less energy when circuit voltage is reduced.

7 BROADCASTING THE MP OVER ROUTED NETWORK

In section 5, we identified different ways of sending the MP. Most networks will be routed in some way. So we first consider if we can broadcast the MP. Figure 1 shows a routed network with two broadcast domains (10.10.40.0 /24 and 172.16.0.0 /16), randomly selected and intentionally kept distant from each other, (Class A and Class B respectively) for ease in configuring the router.

It was possible to send the MP from A to B and C, to shutdown and wake-up. We already knew that
A to B was possible, from results obtained in the previous test; however, here we successfully showed that the MP does route to another domain. Though a single router is used, in a network comprising of many routers, the MP should also be able to achieve the ends of Power Management, provided that routers and firewalls are configured to allow broadcast packets to travel between subnets (Korn et. al., 2006). (The tool used in these experiments is Prof Shutdown 3.2 evaluation version (profshutdown, 2006)).

8 CONCLUSIONS

This paper shows that PM can be supported by MP without detriment to network performance. Our own research indicates that the adoption of PM take-up is at best ad-hoc, sparse and often a small project after business critical objectives have been realised. We believe this to be counterproductive, as buy-in by all stakeholders will itself contribute to the bottom line of businesses and our world. Because the MP piggyback Ethernet, we further believe that the case for its deployment and the potential return on investment (ROI) is significant.

REFERENCES

Companies House (2006) Verbal conversation

Keywords: CTL Kripke model, model update, algorithm, directed graph, implementation.

Abstract: In this paper is presented an update of the Computational Tree Logic (CTL) model checker. The minimal modifications which appear represent the fundamental concept for model the dynamic system. In the paper we use five primitive operations discompose from the operation of a CTL update used already by (Baral, 2005) which presented their approach of knowledge updated on the structures of single agent S5 Kripke. Then we will define the criteria of minimal change for the CTL model update based on these primitive operations. In the final section of this paper are presented the steps of implement the CTL model updated and are described some details of algorithm implementation by applying the model update to the elevator control scenario. The paper (Ding, 2006) is base of results obtained.

1 INTRODUCTION

The verification tools o au tomated formal, such as model checkers, shows delivered diagnosis to provide through automatic error diagnosis in complex designs, examples in (Wing, 1995). The current state of the model checkers technique, as Symbolic Model Verification (SMV) example (Clarke, 2000), Cadence SMV (McMillan, 2002), uses SMV as specification language for both CTL (Computational Tree Logic) and LTL (Lineal Temporal Logic) model checking. Progressing update of the method of the model checkers, begun to employ a formal method for repair approximate error. Since model, checking can handle verification problems complex system and as it may, implemented via fast algorithms, it is quite natural to consider whether we can develop associated algorithms so that they can handle system modification as well. The idea of integrating model checking and automatic modification has been investigate in recent years. In work (Harris, 2003) the model checking is formalized often with an updating operator satisfied the axioms U1-U8 what represent the classical proposition knowledge of updated Katsuno-Mendelzon postulates for belief update (Baral, 2005). They discussed knowledge update and its minimal change, based on modal logic S5. Both the update of the knowledge base and the knowledge update are currently at the theoretical research stage. Their approach of knowledge update could integrate with model checking technology towards a more general automatic system modification. In this paper, we considered the problem of the update of CTL model from both theories. In substance, as the traditional knowledge
based on the update (Winslett, 1990) consider an update of CTL model subdue a principle of minimum inferior change. More, this minimal change are defined be as well to is definite as a process based on of some operational processes which a concrete algorithm for the update of CTL model could be implemented. In the final section of this work, we present a study case where we shown how the system prototype (Ding, 2006) could be applied for the system modified.

2 SYNTAX AND SEMANTICS

CTL is a branching time temporal logic meaning that its formulas interpreted over all paths beginning in a given state of the Kripke structure. A Kripke model $M$ over $AP$ is a tuple $M = (S, R, F)$ where $S$ is a finite set of states, $R\subseteq S\times S$ is a transition relation, $F: S\rightarrow 2^{AP}$ is a function that assigns each state with a set of atomic proposition.

Syntax definition of a CTL model checker (Huth, 2000). A CTL has the following syntax given in Backus near form: $f ::\top | \neg f | f_1\lor f_2 | f_1\land f_2 | f_1\rightarrow f_2 | AX f_1 | EX f_1 | AG f_1 | EG f_1 | AF f_1 | EF f_1 | A[f_1] | E[f_1]$.

A path in $\pi = [s_0, s_1, \ldots, s_i, \ldots, s_j, \ldots]$ such that $\pi = [s_0, s_1, \ldots, s_i, s_j, \ldots]$ and $i < j$, we say that $s_i$ is a state earlier than $s_j$ in $\pi$ as $s_i < s_j$.

Semantics definition of a CTL model checker (Huth, 2000). Let $M = (S, R, F)$ be a Kripke model for CTL. Given any $s$ in $S$, we define if a CTL formula $f$ holds in state $s$. We denote this by $(M, s) \models f$. The satisfaction relation $\models$ define by structural induction on all fourteen CTL formulas (Ding, 2006). We assume all the five formulas CTL presented in the contextually as the paths are satisfied. Be a CTL Kripke model which satisfies the CTL formulas and we considered as a model that can be updated satisfying given formulas. The minimal change should define, based on some operational process, a concrete algorithm for CTL model update that can be implemented.

The CTL update definition: Be a CTL Kripke model $M = (S, R, F)$ and a CTL formula $f$. An update of $M = (S, R, F)$ such that $M' = (S', R', F')$ such that $M' = (M, s_0)$, $(M', s_0) \models f$ where $s_0 \in S$. We use $Upd(M, f)$ to denote the result $M'$ and $Upd(M, f) = M$ if $M \models f$.

3 PRIMITIVE OPERATORS

P1. Add an only relation. Given $M = (S, R, F)$, its updated model $M' = (S', R', F')$ is the result of $M$ having only added one new relation. That is, $s' = S$, $F' = F$, and $R' = R \cup \{(s_{add}, s_{add2})\}$ where $(s_{add}, s_{add2}) \in R$ for one pair of $s_{add}, s_{add2} \in S$.

P2. Remove an only relation. Given $M = (S, R, F)$, its updated model $M' = (S', R', F')$ is the result of $M$ having only removed one existing relation. That is, $S' = S$, $F' = F$, and $R' = R \setminus \{(s_{rem}, s_{rem2})\}$ where $(s_{rem}, s_{rem2}) \in R$ for one pair of $s_{rem}, s_{rem2} \in S$.

P3. Substitute a state and its associated with an only relations. Given $M = (S, R, F)$, its updated model $M' = (S', R', F')$ is the result of $M$ having only substituted one existing state and its associated relations. That is, $S' = S$, $F' = F$, and $R' = R \setminus \{(s_{subst}, s_{subst2})\}$ where $(s_{subst}, s_{subst2}) \in R$ for one pair of $s_{subst}, s_{subst2} \in S$.

P4. Add a state and it associated with an only relations. Given $M = (S, R, F)$, its updated model $M' = (S', R', F')$ is the result of $M$ having only added one new state and it associated relations. That is, $S' = S \cup \{s'\}$, $R' = R \cup \{(s_{add}, s_{add2})\}$ for some $s_{add}, s_{add2} \in S$ and $R' = R \cup \{(s_{rem}, s_{rem2})\}$ for some $s_{rem}, s_{rem2} \in S$.

P5. Remove a state and it associated with an only relations. Given $M = (S, R, F)$, its updated model $M' = (S', R', F')$ is the result of $M$ having only removed one state and its associated relations. That is, $S' = S \setminus \{s\}$, $R' = R \setminus \{(s_{rem}, s_{rem2})\}$ for some $s_{rem}, s_{rem2} \in S$.

All the changes on CTL model can be in terms of all five operations. It can be arguing $P_3$ can be defined in terms of $P_4$ and $P_5$. Anyway, we treat state substitution differently from a combination of state addition and state removed. That is the context, whenever it substitutes a state needed, applied $P_3$ directly more than $P_4$ followed of $P_5$. This thing will simplify definition of minimal change of the CTL model.

For defined the criteria of minimal change of
update CTL model, it needs to consider the changes for both states and relations for the underlying CTL models. We achieve these specifying the differences among states and relations on the models CTL using the primitive operations. Be any two sets X and Y, symmetrical difference among X and Y be denoted as Diff(X, Y) = (X \cap Y) \cup (Y \cap X). Be two CTL models, M = (S, R, F) and M' = (S', R', F') for each primitive operation Pi with i = 1, \ldots, 5, Diff P_i(M, M') indicates the differences between one of two the CTL models where M' is a resulting model from M, that make clear this difference between these operations the types may occur. Since P_1 and P_2 only changes relations, we define DiffP_i(M, M') = (R - R') \cup (R' - R) where i = 1, 2. For the operations P_3, P_4 and P_5, we define DiffP_i(M, M') = (S - S') \cup (S' - S) with i=3,4,5. Although any state changes caused by P_3, P_4, P_5 will imply also correspondence changes on relations, we only count the modifications states and take the state change as the primitive factor in order to measure difference between and M'. For the operations P_i, we should consider the case, which a state is substitute with a new state. For this is necessary difference between these two states to be minimal before the condition of formulated update. A formal algorithm for the proposed CTL model update approach is described in (Ding, 2006) and (Cacovean, 2007).

4 ELEVATOR EXAMPLE

In this section we present a study of case where it is illustrated the features of CTL model updated approaches.

As example, we shall present a scenario for an elevator control system. The designer analyzes the state-transition diagram for the only control transformation, Elevator Controller (EC), finds eight locked-state events (Gomma, 1993). These locked-state events occur because the EC, in most instances, takes one action and then awaits a response before only arrive when the scheduler schedules an elevator. The remaining events can only arrive when the EC is expecting them.

We assume that we have an elevator system control which including in first case, a process for normal moving of lift cabin and in second case, for a faulty process. In first case for the normal moving the elevator cabin process don’t appear with errors, so the door is closed and the passenger going up or down when the button is pressed. For the second process, the faulty process appears when the lift cabin isn’t moving when the button is pressed for start the moving. The aim of the model is where the faulty process appears. The objective of model updating, on other word, is to correct the original model, which contains the faulty process. Starting from the original CTL structure for our propose EC system presented in the figure 1 with eight states denoted with s_1, s_2, \ldots, s_7 and s_d state we added for checking if the elevator is required of another passenger.

The Kripke model has eight states and the propositional variables are from the set \{Start, Close, Move, Error\}. Start (St) represented the start button for start moving up or down the elevator, Close (Cl) represent the close door to the lift cabin, Move (Mv) is moving up or down the elevator and Error (Er) means occur some error.

The formal definition of the Kripke structure of EC is given by M=(S, R, F), where S={s_1, s_2, \ldots, s_7}, R={(s_1, s_2), (s_2, s_3), (s_3, s_2), (s_1, s_3), (s_4, s_5), (s_5, s_6), (s_6, s_7), (s_7, s_6), (s_6, s_4), (s_4, s_5), (s_5, s_4), (s_4, s_3), (s_3, s_4), (s_4, s_d)}, AP={St, Cl, Mv, Er}. The F assigns state s_1 in M with not start, not close, not move and not error, write this as \{\neg St, \neg Cl, \neg Mv, \neg Er\}. State s_2 = \{\neg St, \neg Cl, \neg Mv, \neg Er\}, s_3 = \{St, Cl, \neg Mv, \neg Er\}, s_4 = \{\neg St, \neg Cl, \neg Mv, \neg Er\}, s_5 = \{St, Cl, \neg Mv, \neg Er\}, s_6 = \{St, Cl, Mv, \neg Er\} and s_7 = \{\neg St, Cl, Mv, \neg Er\}. The model shown herebefore:

![Figure 1: The CTL structure of Elevator Controller.](image)

In figure 1 START represented the start elevator, Open and Close represent the open door and close the door, RESET is for a new initialization and
DONE represents the done moving of elevator.

The faulty process from this graph is the path \([s_1, s_2, s_3, s_4]\). The interpretation is: start elevator \([s_1, s_2]\).

In the state \(s_2\) we observed that have not close, that is the door and it isn’t close, and the moving is out of order and it pointed some error. Passed from the state \(s_2\) in the state \(s_3\) where the door elevator shall be close. In the state \(s_3\) has error and the movement of elevator don’t start so it shall push the reset button for the reestablishment. That is, from \(s_3\) passed to the state \(s_4\). Observed that the process with normal move in the case view from the original CTL Kripke structure through \([s_1, s_2, s_3, s_4]\).

Noticed that this model do not satisfies the property \(f = \neg EF(S(I \land EG -Mv))\) (Harris, 2003). The CTL model updated brings a minimum modification of the Kripke model which satisfies the property \(f\). Firstly, it should analyze \(f\) in \(AG(\neg(Str - EG -Mv))\) for remove the symbol \(\neg\). The translation is doing with the function \(Upd_{\neg}\). Then is necessary to check each state whether it satisfies \(\neg(Str - EG -Mv)\). This string shall be parsing before it is checked. Selecting the \(EG -Mv\) to elevator through the model checking function for \(EG\).

In this model, any path has any state when \(\neg Mv\) is selected. Here are searched the paths in the form \([s_1, s_2, s_3, s_4, s_5, s_6, s_7]\) and \([s_1, s_2, s_3, s_4, s_5, s_6]\) which represent the connected components loops satisfy \(EG -Mv\). Then are identified all states with \(St\), these are \([s_2, s_3, s_4, s_5, s_6, s_7]\).

Then are selected the states with \(St\) and \(\neg Mv\), these are \([s_2, s_3]\). Because the \(AG(\neg(Str - EG - Mv))\) formula identifies the model don’t have the both states \(St\) and \(\neg Mv\), is necessary an execution with states \(s_2\) and \(s_3\) so it should apply the updated model. From execution of \(Upd_{\neg Mv}\) function, we shown the case in which applying \(P_1\) on the state \(s_2\) and \(s_3\). The first translate will be from \(\neg(Str - EG - Mv)\) to \(\neg(Str - EG - Mv)\), therefore \(s_2\) and \(s_3\) are updated with any \(\neg St\) or \(\neg EG - Mv\) by the main function \(CTL_{Upd}\) what is dealt with \(\neg\) and with the \(Upd_\neg\) function. In other words, the new states of \(s_2\) and \(s_3\) shall be denoting with \(s_2'\) and \(s_3'\). The \(Upd_{\neg Mv}(M, -\neg(Str - EG - Mv))\) function calls the main function \(CTL_{Upd}(M, -\neg St)\) or \(CTL_{Upd}(M, -\neg EG - Mv)\) for the case \(f_1 \lor f_2\). We choose the \(-St\) because this is simplest than \(-EG - Mv\). In this case is necessary to update the \(St\) in states \(s_2\) and \(s_3\) of path \(\pi\) with \(-St\) instead, then no states on path \(\pi\) have the specification \(EF(Str - EG - Mv)\). The state \(s_2'\) is set \(\{\neg St, Cl, -Mv, Er\}\)

and the state \(s_3'\) is set \(\{\neg St, Cl, -Mv, Er\}\)

The algorithm will generate one of the three resulting models without specific indication, because criteria used are satisfying all the minimally changes from the original model. We consider that our elevator model propose is a model much more simple for understandable and for implemented, because we used a steps method to illustrate this elevator controller. In our case we used the CTL model checker update, verifying all five properties mentioned above which are accomplished also in our case of study.

5 CONCLUSIONS

In this paper, we presented a formal approach for the update the CTL models. Specification of five primitives on the CTL Kripke models (Ding, 2006), define the minimal change criteria of the CTL model updated. Also in this paper are presented semantics and the computing property of approach that we used. The proposed case study is an update principle of minimal change with maximal reachable states, which can significantly improve the update results in modification scenarios of complex system.

REFERENCES


Cacevean L., Popa E.M., Brumar C.I., 2007, Implementation of CTL Model Checker Update, in Proc. 11th WSEAS Int. Conf., COMPUTERS, Greece


Ding Y., Yan Zhang, 2006, ”CTL Model Update: Semantics, Computations and Implementation”. ECAI, Italy.


Wing J. and M. Vaziri-Farahani, 1995, “A case study in model checking software”, in Proc. 3 ACM SIGSOFT.
COMMON TEXTILE VOCABULARIES AND DOCUMENTS
A Conceptual Foundation of a Globally Interoperable Textile e-Marketplace

Jingzhi Guo and Zhuo Hu
Department of Computer and Information Science, University of Macau, Av. Padre Tomás, Pereira, S.J., Taipa, Macau
jzguo@umac.mo, ma66571@umac.mo

Keywords: Textile electronic marketplace, business vocabulary, business document, product data integration, vocabulary integration, document integration, semantic consistency maintenance, semantic interoperability, electronic commerce, electronic business.

Abstract: This paper has proposed a novel common textile vocabulary and document framework (TexVDF) in a collaborative network to enable cross-domain level business information sharing and business document exchange in a semantically consistent way. The approach to this framework is motivated through presenting some real-world examples of business inquiries with product specifications. By these examples, two problems are detected on how to achieve semantic commonality between cross-domain level business vocabularies for textile e-Marketplace mediators and how to allow specificity of cross-domain level common business document templates for local textile e-Marketplace mediators yet still maintaining semantic consistency. To solve these two problems, this paper has firstly reviewed CONEX technologies relevant to the newly developed TexVDF approach, which includes a TexVDF framework, a P2P collaborative textile concept mapping model and a textile business document template model. These two models have been demonstrated by examples to see how they should work.

1 INTRODUCTION

In global textile market, many textile firms face great challenges of increasing global competition to cope with quicker responsiveness of competitors (Teng et al. 2006), better global partnership, and advancing technology for manufacturing and dynamic consumer demand (Ostic 1997). To meet these challenges, textile industry needs to build global e-Markets, where producer, supplier, and retailers can efficiently communicate and exchange information to strengthen their competitiveness. However, textile industry is complex. It involves raw materials such as cotton, silk and polymer that produce both natural and synthetic fibres, which again are converted into many kinds of fabrics and finally to become products such as carpet and apparel.

Besides, textile industry has thousands of large, medium and small sized retailers and manufactures (Teng et al. 2006). They vigorously trade with each other. This indicates that a global e-Marketplace in design must be able to manage the flow of millions of textile products between a very large number of firms. This requires the information exchanged by computers be understandable by all textile participants.

Making textile information understandable between participants on e-Marketplace is an important information interoperability issue (Guo 2007). It requires building a globally interoperable textile e-Marketplace by integrating heterogeneous textile information systems of all participated firms. This at least involves two aspects: the integration of business vocabulary used by all textile firms and the integration of exchanged business documents such as inquiries, offers, counteroffers and orders.

This paper aims to propose a novel common textile vocabulary and document framework (TexVDF) to semantically integrate complex textile firms to enable them to participate in textile e-Marketplace, and also to facilitate the proposed approach as the foundation of the future design of globally interoperable textile e-Marketplace.

The rest of the paper will be arranged as follows: Section 2 will provide a motivational scenario to raise our discussion issues. Section 3 will briefly introduce the relevant technologies and propose a new TexVDF approach to lay a solid conceptual foundation for future design of textile e-Marketplace. Section 4 will exemplify TexVDF approach. Related work is discussed in Section 5. Finally, the conclusion with a contribution list of this paper is given, together with the required future work.
2 A MOTIVATIONAL SCENARIO

e-Marketplace has four properties: distribution, autonomy, interdependence and emergence (Guo 2007). These also apply to textile e-Marketplaces. However, textile e-Marketplaces have a more specific property of *levelled-domain interdependence*, i.e. one level of textile e-Marketplace vocabularies and documents are tightly-coupled with another level of textile e-Marketplace vocabularies and documents. Since they belong to different domains, they have specific requirements. For example, they involve interdependence between levels of industries of raw supplies (e.g. cotton, silk and polymer), fibre, yarn, thread, fabric, printing and dyeing, and garment.

To motivate the problem to solve, Table 1 to 5 summarize the inquiries created by different levels of textile industries to show their interdependence.

Table 1: Raw Cotton Inquiry from Material Industry.

<table>
<thead>
<tr>
<th>Concept</th>
<th>Concept Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Product name</td>
<td>Indian raw cotton</td>
</tr>
<tr>
<td>Type</td>
<td>Shankar-6</td>
</tr>
<tr>
<td>Staple length</td>
<td>29 mm OR 28.5 mm</td>
</tr>
<tr>
<td>Strength</td>
<td>29 gpt or 28 GPT minimum</td>
</tr>
<tr>
<td>MIC</td>
<td>3.5-4.9</td>
</tr>
</tbody>
</table>

Table 2: Polyester Fibre Inquiry from Fibre Industry.

<table>
<thead>
<tr>
<th>Concept</th>
<th>Concept Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Name</td>
<td>Polyester Stable Fibre</td>
</tr>
<tr>
<td>Colour</td>
<td>Semi-dull raw white</td>
</tr>
<tr>
<td>Actual Denier</td>
<td>1.40 + / -0.05 (DE)</td>
</tr>
<tr>
<td>Dry Tenacity</td>
<td>6.50 + / -0.05 (G/D)</td>
</tr>
<tr>
<td>Dry Elongation</td>
<td>30. + / -1.5</td>
</tr>
<tr>
<td>Crimp number</td>
<td>13.00 + / -3.00 EA / Inch</td>
</tr>
<tr>
<td>Degree of Crimp</td>
<td>13.00 + / -0.39</td>
</tr>
<tr>
<td>Shrinkage</td>
<td>7 + / -0.07</td>
</tr>
</tbody>
</table>

Table 3: Blended Yarn Inquiry from Yarn Industry.

<table>
<thead>
<tr>
<th>Concept</th>
<th>Concept Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Name</td>
<td>Yarn</td>
</tr>
<tr>
<td>Colour</td>
<td>white</td>
</tr>
<tr>
<td>Type</td>
<td>Carded</td>
</tr>
<tr>
<td>Composition</td>
<td>Cotton 60%</td>
</tr>
<tr>
<td></td>
<td>Polyester 40%</td>
</tr>
<tr>
<td>Count</td>
<td>32 s</td>
</tr>
<tr>
<td>Quality</td>
<td>high</td>
</tr>
<tr>
<td>Application</td>
<td>Quality knitting and weaving fabrics</td>
</tr>
</tbody>
</table>

An analysis to the above Tables reveals at least two of the following problems:

1. Term usages are not only industry domain-specific. The higher-level industry has a high frequency to use the vocabularies of lower-level industry, but practically each level of industry has its own explanations on their terms in vocabularies.

This problem can be abstracted as a research issue of *cross-domain level business vocabulary commonality*, which states that all levels of textile industry shall be able to semantically communicate with each other via a common vocabulary model.

Table 4: Dyed Fabric Inquiry from Fabric Industry.

<table>
<thead>
<tr>
<th>Concept</th>
<th>Concept Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Name</td>
<td>Fabric</td>
</tr>
<tr>
<td>Type</td>
<td>Combed</td>
</tr>
<tr>
<td>Colour</td>
<td>Yarn dyed</td>
</tr>
<tr>
<td>Technique</td>
<td>woven</td>
</tr>
<tr>
<td>Style</td>
<td>Jacquard</td>
</tr>
<tr>
<td>Usage</td>
<td>Garment</td>
</tr>
<tr>
<td>Composition</td>
<td>Cotton 60%</td>
</tr>
<tr>
<td></td>
<td>Polyester 40%</td>
</tr>
<tr>
<td>Construction</td>
<td>45 s x 45 s</td>
</tr>
<tr>
<td>Density</td>
<td>130 x 70</td>
</tr>
<tr>
<td>Width</td>
<td>57/58”</td>
</tr>
</tbody>
</table>

Table 5: Men’s Coat Inquiry from Garment Industry.

<table>
<thead>
<tr>
<th>Concept</th>
<th>Concept Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Name</td>
<td>Coat</td>
</tr>
<tr>
<td>Fabric</td>
<td>Fabric</td>
</tr>
<tr>
<td>Cotton</td>
<td>60%, ring, spun</td>
</tr>
<tr>
<td>Polyester</td>
<td>60%</td>
</tr>
<tr>
<td>Weight</td>
<td>300 grams/square meter</td>
</tr>
<tr>
<td>Specification</td>
<td></td>
</tr>
<tr>
<td>Zipper</td>
<td>Front</td>
</tr>
<tr>
<td>Pocket</td>
<td>Pouch</td>
</tr>
<tr>
<td>Cuff</td>
<td>rib, lycra (cotton 60%, polyester 40%)</td>
</tr>
<tr>
<td>Bottom</td>
<td>rib, lycra (cotton 60%, polyester 40%)</td>
</tr>
<tr>
<td>Size</td>
<td>S, M, L, LL, LLL, LLLL</td>
</tr>
<tr>
<td>Drawcord</td>
<td>Fabric</td>
</tr>
<tr>
<td>Use</td>
<td>Men</td>
</tr>
</tbody>
</table>

This problem can be abstracted as a research issue of *cross-domain level business document commonality*, which states that all levels of textile industry shall be able to personalize document templates from common document templates.

3 TEXVDF APPROACH

In this section, we will propose a novel common Textile Vocabulary and Document Framework (TexVDF) to solve the above two problems to lay a
solid foundation for future globally interoperable textile e-Marketplace. We vision that this e-Marketplace will be a common textile information space that all kinds of services and collaborative activities will be enabled, based on our designed TexVDF solution.

In the following, we will first discuss the relevant technologies and then propose the solution framework and its details.

### 3.1 Relevant Technologies

In CONEX research (Guo 2008), a generic e-Marketplace is designed like a multi-sons solar system as a set of common collaborative service mediators, each having a set of local collaborative service designers for their own service users on CONEX Network (ConexNet). It can be shown in Figure 1.

In this CONEX e-Marketplace, business vocabularies and documents are firstly collaboratively designed at mediators (M) on a P2P collaborative network. Mediators include designer roles of common vocabulary design, dictionary design and document template design. These designers are knowledge experts (or knowledge engineers) and can make sure that things go right. Thus, they are called dominators (D) when their results are sold to local firms. The local firms subscribing or purchasing mediators' design services localizes these services into their own personalized forms, that is, local business vocabularies and document templates. The local designers are knowledge workers but are not experts. Thus, they are followers (F) of the dominators (D). The collaboration between dominators and followers is a dominator-to-follower (D2F) relationship on a point-to-point communication network. The follower cannot modify the common design but follow to generate their own. When local designers have designed local information about firm-based vocabulary and document templates, their users of their own firms can then automatically exchange business information, such as inquiries and offers.

In CONEX e-Marketplace, vocabularies and documents are designed following Product Map (PM) theory (Guo 2008). Its implementation is specified in CONEX Grammar (Guo 2008).

The information exchange by PM from one local firm (L) to another local firm (L) follows a concept supply chain (Guo 2008), such that:

- Concept(L1) map onto Concept(M1) map onto Concept(M2) map onto Concept(L2),

where concept(L1) of L1 finally arrive at L2 as concept(L2).

Since concepts of L1, M1, M2 and L2 are all collaboratively created and mapped, there is no semantic inconsistency between them in theory. Thus, CONEX e-Marketplace provides an approach of accurate yet automatic information exchange in the perspectives of all users (U) of L.

### 3.2 TexVDF Framework

The TexVDF framework follows the design of CONEX e-Marketplace but adds the new layer thought to the framework. To provide a smooth discussion, we illustrate this framework in Figure 2, where different levels of textile industry have been separated but integrated in a coherent collaboration framework.

In Figure 2, CONEX e-Marketplace has been extended to include six layers of textile e-Marketplaces, which are:

- **Raw material e-Marketplace**, which focuses on trading cotton, silk, wool, fur, feather, plant, polyester, polypropylene, polyethylene, etc.
- **Fibre e-Marketplace**, which focuses on trading fibres made from cotton, silk, polymer (e.g. polyester, Dacron, nylon, Micron), plant (e.g. corn, soybean), etc.
- **Yarn and thread e-Marketplace**, which focuses on trading yarns of cotton, polyester, blended polyester, wool, plant, different threads, etc.
- **Fabric e-Marketplace**, which focuses on trading woven and non-woven fabrics of cotton, polyester, blended polyester, metallic, etc.
- **Printing and Dyeing e-Marketplace**, which focuses on trading dye, dyeing services of various yarns, threads, fabrics, printing, etc.
- **Garment e-Marketplace**, which focuses on finished textile products like jacket, pullover, shirt, protective garment, gown, gloves, etc.

By this layered classification of textile e-Marketplaces, a new technology of P2P collaborative mapping between e-Marketplace mediators is developed to solve the problems stated in Section 2. This technology complements the previous CONEX technology to enable cross-domain level vocabulary and document interoperability.
A standard terminology servicing centre is added to enable the adoption of existing standard terms such as standard measure, currency code, and de facto standard use of textile terms. This service will reduce the collaborative work effort occurred in both vocabulary and document templates for cross-domain levelled textile e-Marketplaces.

Particularly, the TexVDF framework adopts a collaborative mapping solution to achieve semantic commonality of cross-domain level vocabularies and a specificity solution to using common textile document template.

In the next two subsections, we will elaborate the design of these two solutions.

### 3.3 P2P Collaborative Mapping

A **P2P collaborative mapping model** (CMM) is a solution to enabling various mediators of levelled textile e-Marketplaces to collaboratively work together to map their textile concepts in a common level but with possible expression variations for their own. This model is necessary because there are many synonyms and homonyms appeared in textile industry. Each textile mediator has its own concept mediation context and is not possible to define all synonyms and homonyms for each textile concept. The impossibility for each textile mediator to include all is due to the complexity that textile terms are developed from different contexts, where synonyms and homonyms come from various textile types, usages of dated and obsolete terms, term derivations, languages of origin, and toponyms (Leech 1999).

For example, “baby combing wool” has variation of “French combing wool”; “linen lawn” is synonymous with “handkerchief linen”; and “drab” is a homonym with meanings of “colour of a moderate to grayish or light grayish yellowish brown or light olive brown” and “cloth of a light dull brown or grayish brown or unbleached natural colour, especially a heavy woollen or cotton fabric”.

To bind the concept expressions of individual mediators with a common concept expression, the mediators’ concept expressions must be framed in a common structure, where they can freely express their ideas of categorizing their individual textile concepts. CCM provides this function in the following form, such that:

**Definition 1 (CMM). Collaborative Mapping Model**

Given a set of individual concept expressions $\sigma_1, \sigma_2, \ldots, \sigma_n$ and uniquely identified by iid$_1$, iid$_2$, ..., iid$_n$ under different mediators’ contexts $x_1, x_2, \ldots, x_n$ about a piece of common formal concept $FC$, defined by AN, uniquely identified by IID, typed by CT, inherited from IF, concept or context of CO, and possibly referenced to (RT) a standard term of a terminology, such that:

$$\Gamma \models \text{concept}[\text{IID}, \text{FC}, \text{AN}, \text{CT}, \text{IF}, \text{CO}, \text{RT}] \Rightarrow \text{map}[\text{concept}[\text{IID}, \text{iid}_i, \sigma_i]@x_i], \ldots, \text{concept}[\text{IID}, \text{iid}_n, \sigma_n]@x_n].$$

where, “$\Gamma$” is a P2P collaborative mapping engine that enables all mediators to work together in real-time, “|$\models$” is a declaration that all mapping follow the mapping structure that is defined, and “$\Rightarrow$” refers to that the semantics of each mediator’s concept strictly follows the semantics of common concept for all.

The above CMM has following principles:
(1) Each mediator’s semantics about the same concept must be semantically consistent by P2P collaboration following CMM structure.

(2) Personalized concept expressions of individual mediators, rather than the formal concept FC, are allowed to adapt to the local mediator’s e-Marketplace environment but have to be mapped onto common concept via IID.

(3) Standard terms of terminology, from standard terminology servicing centre shown in Figure 1, are encouraged to be referenced during common concept design.

Particular to CMM, some details are important to mention.

Definition of annotation AN. It is a full definition of a common concept, and not a single word or phrase that may not fully capture the meaning of the concept, or may lead to sense ambiguity.

Formal concept FC. It is a machine-readable term about a common concept. It may capture full or almost full meaning of the concept definition AN, but not guaranteed for accuracy. It is primarily used for information retrieval for search services. FC is typed as a set to include multiple words and phrases with exact or similar meanings to the defined concept such as abbreviation.

Internal unique concept identifier IID. It is unique identifier of AN with semantic causal order relationship such as AN ⇒ IID. It implies that any IID cannot be created to use without the meaning of AN conveyed in IID. If such happens, it will be prohibited to process.

Reference to a concept RT. It is a reference to a semantically equivalent concept, often a well-defined term in terminology or an already-defined vocabulary by CMM. It is typed as a namespace use.

Context of a higher level concept CO. It defines the direct context of the current concept in a vocabulary hierarchy. For example, given “domestic appliances (domestic refrigerators)”, the “domestic appliances” is the direct context of “domestic refrigerators”. It is similar to a broad term (BT) in relation to a narrow term (NT) in controlled vocabulary (Fidel 1999). The CO is important for efficient and accurate information exchange. It is also very useful for accurate machine translation by word sense disambiguation (Vickrey et al. 2005) through CO context.

Inheritance from sources IF. It defines that where the concept is inherited or derived. It states the origin of the concept. It is useful for improving information retrieval and disambiguating the sense of the concept when the concept is applied in machine translation. For example, “ramie cotton blended fabric” is inherited from both “ramie” and “cotton” and under the context of (CO) of neither “ramie fabric” nor “cotton fabric” but “blended fabric”. The inherited terms of “ramie” and “cotton” are useful to infer the concept meaning.

Concept type CT. It defines in which term type the concept applies. For example, “bombazine” means “a twilled or corded dress-material, composed of silk and worsted; sometimes also of cotton and worsted, or of worsted alone”. Thus, it belongs to the domains of “silk” and “cotton” as a noun-form, but it also belongs to the large domain of “textile” as an adjective form to refer to “worsted”.

With above descriptions, the common concept concept[IID, FC, AN, CT, IF, CO, RT] is clear for individual mediators to collaboratively map onto their own local forms in map[concept[IID, iid1, σ1]@x1], ..., concept[IID, iidn, σn]@xn. In the local form, practically, the context xi can be designed as an individual mediator’s unique identifier (LID) to refer to its specific context.

3.4 Document Template Model

A textile document template model (Doclate) is a solution to enabling specifying business document templates to a domain-specific level but still being able to utilize the common document templates prepared in advance and to consistently use common concepts designed by CMM.

To realize this model, this paper regards a business document template a set of hierarchical concepts and improves the previous work of CODEX (Guo 2006) by removing P2P collaboration requirement and adding RT to CMM, such that:

Definition 2 (Doclate). Document Template Model

Given a set of concepts defined in CMM, there is a common document template (com) with a set of concepts (called elemon) hierarchically identified as IID, annotated by AN, in the context of CO, presented as DP, referenced to RT of CMM, and have occurrence OC; each “elemon” has a reification structure “value” to reify the concept with representation format PT, data type DT and possible function FN for managing reification. This “com” will further be extended to (⇒) “loc” to provide specificity of “com”, such that:

\[ Γ \rightarrow \text{com: elemon[IID, AN, CO, DP, OC, RT](value[PT, DT, FN])} \Rightarrow \text{loc: elemon[IID, AN, CO, DP, OC, RT](value[PT, DT, FN])} \]

where “Γ” is a business document template editing engine, “⇒” declares that the editing follows the structure defined.

The above model has the following principles:

(1) Each common document template “com” is only a semi-finished template and could be further
specified in any way to “loc”, but both must follow the Doclate structure model.

(2) Both “com” and “loc” can only create “element” concepts through RT to CMM vocabularies.

Particularly, the Doclate structure elements have the following semantics:

AN, IID, CO, and RT. They exactly have the same interpretation as in CMM, but RT refer to CMM.

Document concept display phrase DP. It is a phase for visual display to represent concept. For example, a concept of “an appliance, a cabinet, or a room for storing food or other substances at a low temperature” could be visually displayed as “refrigerator” or “domestic refrigerators” as needed in the different designs of Doclate template.

Concept occurrence OC. It defines occurrence of concept in designed Doclate template. The concept occurrence may happen. For example, “product item” in a purchasing order may occur many times for different purchased items.

Value concept VALUE. This is a reification symbol to introduce a reification of a concept to a particular concept, for example, “colour” → “red”.

Presentation style of reified concept PT. It defines how a reified concept should be displayed. For example, “1” could be displayed as “1”, “one”, or “USD1/piece”. It is a logic module implemented in a remote namespace.

Data type of reified concept DT. It defines the data type of the reified concept, for example, “string” or “decimal”.

Operational function of reified concept FN. It defines how the reified concept value could be computed, for example, automatically generated reified “date” value, reified computational group concept “total”, or a result of a logic module.

With the above descriptions, common document templates could be easily specified locally but accurately maintain semantic consistency between using parties of different levels of textile e-Marketplace.

Table 6: Textile Common Vocabulary for Garment

4 COAT EXAMPLE ON TEXVDF

In this section, we demonstrate TexVDF approach through a garment inquiry specified in Table 5. The following examples will adopt the written form of vector concept tree like “i...i” to represent the concept hierarchy of any vocabulary and document template (Guo 2008).

4.1 CMM for Coat Vocabularies

To implement the garment inquiry example, we first design the common vocabulary, shown in Table 6, based on CMM of definition 1.

With the above collaboratively designed garment vocabulary commonly for all levels of textile e-Marketplace mediators, individual mediators can localize their personalized vocabularies applicable to their own e-Marketplaces. Table 7 shows the localized mapping for one English mediator x1 and a Chinese mediator x2 for their own e-Marketplaces.

Table 7: Mapping of Local Vocabulary onto Common Vocabulary

Given the above local mapping onto the common vocabulary shown in Table 6, the users of local textile e-Marketplace of x1 and x2 can then exchange business information without any semantic conflicts.

4.2 Doclate for Coat Inquiry

In this part, we exemplify the work of Doclate model in the example of Table 8 and Table 9.

Table 8: Common Document Template for Inquiry

```plaintext
<table>
<thead>
<tr>
<th>concept</th>
<th>id</th>
<th>fc</th>
<th>an</th>
<th>dp</th>
<th>rt</th>
</tr>
</thead>
<tbody>
<tr>
<td>cotton</td>
<td>1</td>
<td>&quot;cotton&quot;</td>
<td>&quot;cotton&quot;</td>
<td>dp</td>
<td>coat</td>
</tr>
<tr>
<td>polyester</td>
<td>2</td>
<td>&quot;polyester&quot;</td>
<td>&quot;polyester&quot;</td>
<td>dp</td>
<td>coat</td>
</tr>
<tr>
<td>fabric</td>
<td>3</td>
<td>&quot;fabric&quot;</td>
<td>&quot;fabric&quot;</td>
<td>dp</td>
<td>coat</td>
</tr>
<tr>
<td>size</td>
<td>4</td>
<td>&quot;size&quot;</td>
<td>&quot;size&quot;</td>
<td>dp</td>
<td>coat</td>
</tr>
<tr>
<td>weight</td>
<td>5</td>
<td>&quot;weight&quot;</td>
<td>&quot;weight&quot;</td>
<td>dp</td>
<td>coat</td>
</tr>
<tr>
<td>pocket</td>
<td>6</td>
<td>&quot;pocket&quot;</td>
<td>&quot;pocket&quot;</td>
<td>dp</td>
<td>coat</td>
</tr>
<tr>
<td>cuff</td>
<td>7</td>
<td>&quot;cuff&quot;</td>
<td>&quot;cuff&quot;</td>
<td>dp</td>
<td>coat</td>
</tr>
<tr>
<td>hood</td>
<td>8</td>
<td>&quot;hood&quot;</td>
<td>&quot;hood&quot;</td>
<td>dp</td>
<td>coat</td>
</tr>
<tr>
<td>drawcord</td>
<td>9</td>
<td>&quot;drawcord&quot;</td>
<td>&quot;drawcord&quot;</td>
<td>dp</td>
<td>coat</td>
</tr>
<tr>
<td>zipper</td>
<td>10</td>
<td>&quot;zipper&quot;</td>
<td>&quot;zipper&quot;</td>
<td>dp</td>
<td>coat</td>
</tr>
<tr>
<td>bottom</td>
<td>11</td>
<td>&quot;bottom&quot;</td>
<td>&quot;bottom&quot;</td>
<td>dp</td>
<td>coat</td>
</tr>
<tr>
<td>garment</td>
<td>12</td>
<td>&quot;garment&quot;</td>
<td>&quot;garment&quot;</td>
<td>dp</td>
<td>coat</td>
</tr>
<tr>
<td>width</td>
<td>13</td>
<td>&quot;width&quot;</td>
<td>&quot;width&quot;</td>
<td>dp</td>
<td>coat</td>
</tr>
<tr>
<td>length</td>
<td>14</td>
<td>&quot;length&quot;</td>
<td>&quot;length&quot;</td>
<td>dp</td>
<td>coat</td>
</tr>
<tr>
<td>material</td>
<td>15</td>
<td>&quot;material&quot;</td>
<td>&quot;material&quot;</td>
<td>dp</td>
<td>coat</td>
</tr>
</tbody>
</table>
```
In this simple document template, some content of the garment inquiry sheet has been designed. It then can be further defined by local e-Marketplace designers of document templates. The function of common level document templates is to reduce the time of local e-Marketplace document designers and thus to lower the e-Marketplace cost.

When local e-Marketplace designers obtain the rough document templates, they personalize them as their own needs, as shown in the example Table 9.

Table 9: A Particular Garment Inquiry Sheet Template.

<table>
<thead>
<tr>
<th>Element ID</th>
<th>Attribute</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>an</td>
<td>cotton</td>
</tr>
<tr>
<td>1.1.1</td>
<td>an</td>
<td>fabric</td>
</tr>
<tr>
<td>1.1.2</td>
<td>an</td>
<td>weight</td>
</tr>
<tr>
<td>1.2</td>
<td>an</td>
<td>size</td>
</tr>
<tr>
<td>1.3</td>
<td>an</td>
<td>use for</td>
</tr>
<tr>
<td>1.4</td>
<td>an</td>
<td>specification</td>
</tr>
<tr>
<td>1.4.1</td>
<td>an</td>
<td>zipper</td>
</tr>
<tr>
<td>1.4.2</td>
<td>an</td>
<td>pocket</td>
</tr>
<tr>
<td>1.4.3</td>
<td>an</td>
<td>hood</td>
</tr>
<tr>
<td>1.4.4</td>
<td>an</td>
<td>cuff</td>
</tr>
<tr>
<td>1.4.5</td>
<td>an</td>
<td>drawcord</td>
</tr>
<tr>
<td>1.4.6</td>
<td>an</td>
<td>bottom</td>
</tr>
<tr>
<td>1.4.7</td>
<td>an</td>
<td>zipper</td>
</tr>
</tbody>
</table>

In the above Table, a more specified garment inquiry sheet template has been designed. With this document template, users of local textile e-Marketplace can reify the template by filling the “value” information to automate inquiry exchange.

5 RELATED WORK

Textile e-Marketplace design that enables business information sharing and business document exchange is an important research field. In this design, semantic integration of textile vocabularies and documents are the foundation for a globally interoperable textile e-Marketplace. Currently, active researches can be found in DAMA and Moda-ML.

In U.S., TEXNET is a textile industry data-sharing network to address data sharing among business partners. It presents shared data in the screen or saved it in standard formats on a local platform (Lovejoy, a). Particular to the textile e-Marketplace, DAMA (Chapman et al. 2000) is a project of such type. DAMA applies a pipeline analysis method (Lovejoy, b). In DAMA’s research, supply chain concept is adopted through Supply Chain Integration Program (SCIP), where an inter-enterprise decision support tool is developed to analyze supply chain tradeoffs. DAMA is designed to use TEXNET for data sharing. It intends to support information sharing and decision making between firms of retail, apparel, textile and fibre within a particular supply chain. While supply chain method for textile e-Marketplace integration is worth investigating, it has some entry limitations for small and medium sized enterprises to participate in.

In Europe, MODA-M (MODA-ML; Leech 1999) is a research on textile e-Marketplace. Its approach is based on the exchange of standardized XML documents, where ebXML protocol has been adopted to transfer XML messages. Moda-ML is ontology-based in design. It generates a modular ontology where each basic concept can be managed independently from the others and is identified by its own namespace (Gessa 2007). It has defined a common platform (Gessa et al. 2004), which attempts to be adopted by firms to improve their interoperability. Ontology-based ebXML document exchange is an attractive approach and fits in most research prototypes in many other e-Marketplaces. However, its success depends on the ontology interoperability by its own in semantic level.

It is necessary to make a clear distinction between the three general strategies of integrating heterogeneous business information (Guo 2008). They are mandatory standardization (a standard is enforced for all participants, in which heterogeneous information integration between standard systems and the participants’ local systems is the task of the local participants), automated mediation (an intelligent agent as a mediator to mediate heterogeneous information between disparate participants’ systems based on predefined rules, in which if no rules can be applied, mediation of heterogeneous concepts is not accurate), and collaborative conceptualization (heterogeneous business concepts between disparate local participants’ systems must be collaboratively mapped on a higher level common system before they can be exchanged).

This paper adopts collaborative conceptualization strategy to avoid the weakness of mandatory compliance of standards by users and inaccurate concept mediation by intelligent mediator. Under this strategy, mediator is designed as a collaborative mediator to ensure the semantic consistency between heterogeneous concepts. This strategy absorbs some of the merits from both mandatory standardization and automated mediation strategies, but it creates no standards and mixes automated agents with human.

6 CONCLUSIONS

This paper has proposed a novel common textile vocabulary and document framework (TexVDF) in a collaborative network to enable cross-domain level
business information sharing and business document exchange in a semantically consistent way. The approach to this framework is motivated through presenting some real-world examples of business inquiries with product specifications. By these examples, two problems are detected on how to achieve semantic commonality between cross-domain level business vocabularies for textile e-Marketplace mediators and how to allow specificity of cross-domain level common business document templates for local textile e-Marketplace mediators yet still maintaining semantic consistency. To solve these two problems, this paper has firstly reviewed CONEX technologies. Then, TexVDF approach is presented in a TexVDF framework, which is an improvement of previous CONEX model. To realize TexVDF framework, a P2P collaborative concept mapping model and a textile business document template model have been developed. The former has resolved the problem of semantic commonality of local mediators' individual vocabularies, and the latter has solved the problem of designing common business document templates and their flexible specificity to document templates adaptable to local textile e-Marketplaces. These two solution models have been demonstrated in examples to see how they could work.

The TexVDF approach has been provided as a conceptual foundation for future design of globally interoperable textile e-Marketplaces. It has advantages compared existing solutions. (1) It does not enforce standards on business vocabulary and document templates. This implies a flexible solution to semantic consistency maintenance between participated textile e-Marketplaces. (2) The new framework it provides allows cross-domain level semantic interoperability but still enable personalization. (3) Useful terminology standards are welcome to be flexibly integrated into the new framework. It implies a standard integration but not rigid. (4) Vocabulary design between e-Marketplace mediators is collaborative in real-time. This enables semantic accuracy and avoids erroneous inference between individual e-Marketplaces for those taking the approach of independent vocabulary design. (5) Document template design and specificity adopt a simply hierarchical document structure where each document element concept referenced to a well-defined collaborative concept vocabularies in CMM. This enables simple creation and use of document templates.

Currently, the improved structure specification of TexCVF framework in terms of XPM is in final release stage. The future work of this paper will be the implementation of the textile e-Marketplace based on this conceptual foundation.

ACKNOWLEDGEMENTS

The work reported in this paper has been supported by University of Macau Research Grant.

REFERENCES


MODA-ML. http://www.moda-ml.org


PROPOSAL OF AN ARCHITECTURE FOR DIGITAL CITIES CREATION

Proposal of an Architecture P2P for Digital Cities Creation

André M. Panhan, Denys G. Santos and Leonardo S. Mendes
Communication Networks Laboratory, Communications Department, Electrical and Computer Engineering Faculty
University of Campinas – UNICAMP - Cidade Universitária "Zefirino Vaz"- Barão Geraldo, Campinas, SP - Brazil
apanhan@gmail.com, gsdenys@gmail.com, lmendes@decom.fee.unicamp.br

Keywords: Digital Cities, Middleware, Open Access Metropolitan Networks.

Abstract: The term digital cities do not have a precise definition. The number of digital cities, whose development started from the 90s, is constantly growing. Motivated by the increase of these cities, this paper describes architecture for digital cities environment as a proposal to create a middleware for the interconnection of the various segments of society. This middleware is structured in a Platform for P2P networks, also known as DHT, with emphasis in the JXTA modified protocol.

1 INTRODUCTION

Also known as Virtual City, Digital City, Intelligent City and other names, digital cities represent a kind of projection of a city in the digital world and emerge as one of the forces that may contribute to space organization (Silvia, 2002).

The digital cities can have many definitions. In (Silvia, 2002), it is defined that digital cities are people, systems and institutions that remain connected through a digital communication infrastructure. In digital cities, one uses real cities as reference for their digital counterpart, aiming at developing social relations of different purposes. These include one or more of the items cited below.

- Creation of political and cultural spaces for people and groups manifestation;
- Creation of communication channels between individuals and groups;
- Creation of communication and negotiation channels between municipal government and citizens;
- Favoring a greater identification from the visitors with the reference city;
- Creation of a collection of information from many different species and easily accessible on the reference city.

Digital Cities, as defined by Graham (Graham, 2007), can serve as a tool to improve communication between citizens and local government, stimulating many activities that promote new opportunities to the citizens, as we identify below.

"... virtual cities are electronic spaces, generally based on the World Wide Web, which were developed to connect, in an explicit way, the development schedules of each city. These virtual cities are working as political tool for a variety of urban plans and objectives: global urban marketing, stimulus to business and consumption, improving communications between citizens and local governments, increase the local enterprise competitiveness, greater local economies integration and the rebirth of civic and local culture." (Graham, 2007)

Motivated by the increased number of digital cities developed during the 90s, this article presents an architecture for digital cities environments and a proposal to create a middleware based on a DHT structure to people, groups and services (public and private) interconnections for this proposed environment.
2 DIGITAL CITIES ARCHITECTURE

In (Yamaoka, 2006) it is defined a minimal set of assumptions, policies and technical specifications that rules the use of Information and Communication Technologies (ICT) by the government, establishing the conditions for interaction. This middleware emphasizes the JXTA modified protocol, extending the interoperability traditional models in government for other society segments as citizens, companies and the third sector. Our proposition is developed based on these minimal premises.

The proposed architecture establishes rules on the following areas responsible for the interoperability of various society segments, Interconnection, Security, Access Media and organization and Information exchange.

For each of these areas we specify components, for which standards are established.

2.1 Interconnection

The segment "Interconnection" establishes the rules for enterprises (such as service or commercial companies) and governments to interconnect themselves upon the local network, which can be a metropolitan network, a community network, or even the Internet. This segment also treats the interoperability rules of enterprises and governments with the citizen. In this segment are defined Electronics Message Transport, Electronics Message Safely Content, Intercommunication LAN/ WAN and Transportation.

2.2 Security

This segment deals with the proposed architecture ICT security aspects that should be considered for Digital Cities creation. We treat in this segment security IP patterns, encryption, System Development and Network Services.

2.3 Access Media

In the segment Access Media one explains issues relating to devices standards, which are used to access the services offered by Digital Cities. We also define how devices must be treated. Some of these devices are computers, mobile phones, hand-holds and digital television, for example.

2.4 Organization and Information Exchange

This segment deals with aspects related to the processing and transfer of information among services offered by Digital Cities. It includes components for data exchange, data processing, data definition and Standards Database Catalogue (SDC). Information distribution for these situations requires the construction of suitable metadata structures for each of the cases above.

Clearly defined policies and specifications for interoperability and information management are key issues for providing the enterprise connection among citizen, enterprise, government and service providers. The Digital City is designed as a basic structure for the government strategy and business in the several business scenarios (B2B, B2C, B2G, G2G, G2C and C2C). This approach allows the rationalization of investments in ICT, through the sharing, reuse and exchange of technology resources.

Enterprises, government and services providers information resources are valuable economic assets. By ensuring that information or service can be quickly located and exchanged between the public and private sectors, keeping the privacy obligations and security, one encourages and stimulates local economy.

Nowadays, most of the countries of world recognize the needs for high investments in developing universal access to multiservice networks for all citizens. Economic analyses show that universal services cause a direct and almost immediate increase in GNP. Its also understood that universal access to services, particularly when we talk about e-Gov services, can only be achieved through the creation of virtual (and, sometimes, real) local networks (or Digital Cities). In creating the Digital Cities one is primarily concerned with establishing policies development, procedures and standards for ICT, mounting dedicated structures to reach local interoperability and seeking better quality provision services at reduced costs. The existence of a middleware as a foundation for
electronic services creation is a pre-requisite for better services provision at a reduced cost for all the actors of the Digital City.

3 MIDDLEWARE

To create the Middleware, we propose to add an abstract slice on JXTA architecture. This slice will be acting as an access control to digital cities network.

3.1 Structure

The basic structure to the proposed middleware is based on the JXTA overlay network topology as can be seen in Théodoloz (2004), or in Traversat (2003). This model creates a network of Rendezvous servers to establish interconnection links among these, and forming a complete graph, where each server is a vertex and the connections are the edges.

In a JXTA network each server has a complete and updated list of connected peers. Therefore, it is possible that the search for a service in the network be performed in a linear order when the network servers are consistent. Otherwise, when an inconsistency is found in the server’s network, another mechanism must be activated to return the identifier of a peer in time $n$.

Figure 3 represent the proposed architecture for middleware model described in this paper. In this architecture the $Rx$ peers provide routing, the $Ex$ peers services and the $Px$ peers access to the interaction portal of Digital City.

Another relevant element to build a Digital City is the creation of a protocol of multiple connections to the peers connected to the network. In order to use the multi-connection resources already available in the JXTA architecture and seeking to make an efficient use of the network and maintain its workload low, it is proposed the creation of an abstraction layer that keeps the peer connected to only one Digital City. However, the protocol allows the peer to make connections to other peers and other Digital Cities without loosing the main connection. This resource allows that the peer uses services from other digital cities and, at the same time, allows other digital cities to access its services.

To provide this resource, the abstraction layer cannot treat the peer identifier like it is treated in the JXTA architecture. In the proposed abstraction layer, the peer identification must be done using a pair of identifiers (GPID, PRID), where GPID is the group identifier and PRID is the peer identifier.

3.2 Security

In a corporative or government communication environment, data security is an extremely important issue both in storage and transport of information.

To provide security in the operations of data applications and transactions, one can use resources offered by the JXTA API. This resource has a set of algorithms that provide a base to security control in the applications developed using this API.
3.3 Digital City Directory

To maintain a consistent identifier services list in the digital cities network, one can create agents associated to the digital city portals and systems interfaces. These agents can perform periodic searches in the network to look for active services and to map these services with an appropriate description. With this map, the creation, ordering and availability of data are responsibility of the respective digital city through their portals and access interfaces.

4 USE CASE

The concept of Digital City can be easily explained how a user makes use of the digital city environment to plan and contract a travel to a congress in another city.

The citizen enters in the portal a make authentication. Authentication done, the citizen makes them registration and makes a payment with a credit card. As the bank system also was developed within digital city architecture, the congress page makes a request to the bank system with totally transparent to the user.

After registration, the citizen chooses a hotel in the city of the congress, buy the airplanes tickets and rent a car, paying all of this in the same bank systems.

Although these actions performed by the citizen are familiar to the common user of the Internet, the Internet users are not able to follow, and even interfere in the process. In the Digital City, the user is able to perform all these activities by accessing only one page and yet follows all the process that happening behind the scenes.

5 CONCLUSIONS

The creation of digital cities provides communication channels to facilitate and improve the relations between citizens and government, citizen and enterprises, and enterprises and government. This communication environment is suitable to head these relationships because the sophisticated identification, authentication and security offered by the digital city middleware. For the authenticated user, the digital city can provide an environment rich in information and services.

The DHTs proposed in this work has as one of its main characteristics the separation between identifier and locator. This feature provides superior mobility and improves security. Another important feature of this DHT is the service doesn’t have mapped in a physical address. The service can be mapped in a simple name, allowing an improved transparency with respect to migration.

To provide interoperability among different digital cities without creating inconsistency or service address replication, the creation of the abstraction layer proposed in this paper is indispensable.

ACKNOWLEDGEMENTS

The work presented here has been developed under the umbrella of the projects “Municipal Infovia – An Open Access Network for Cities” and “SIGM – An Integrated e-Gov Environment for Cities”. These projects have been supported in part by the governments of the cities of São José do Rio Preto, Pedreira, Penápolis, and Campinas, São Paulo State, Brazil.

REFERENCES


Keywords: e-Business, Open Source Software, Small-to-Medium Enterprise (SME).

Abstract: This research investigated the issues that impact on an SME adopting e-Business. It was found that many barriers may exist for e-Business adoption in an SME. One of which is limited financial resources to purchase internet technology required for e-Business adoption and development. However, open source technology has grown in popularity for a number of years, with governments, business firms, military and educational organisations incorporating open source software. This paper provides an overview on open source technology which could be adopted for e-Business architecture for the development of e-Business applications. Furthermore, it provides two action research case studies on SMEs adopting open source software for the development of e-Business capabilities. It was found the projects were successfully developed and implemented for each case study company. Although successful, the research team found that more research is required in open source software for the development of e-Business applications.

1 INTRODUCTION

In the United Kingdom (UK), Small-to-Medium Enterprise, (SMEs) provides a significant contribution to the economy. Some authors report the contributions to Gross Domestic Product (GDP) up to 40 percent (Taylor and Murphy, 2004). The world economy is changing, with growth in emerging markets and developing economies continuing to outstrip that in developed economies, (Jutla et al, 2002). This has resulted in more pressure on SMEs to discover ways in which to deliver its products and services more effectively, and improve operations activities to remain competitive.

The growth in the internet and e-Business applications has provided organisations with new opportunities to offer products and services, share information with business partners and meet customer demands. Therefore, e-Business applications and the internet have the potential to increase the competitiveness and growth of small firms, (Paliwoda, 2004). E-Business applications provide an organisation with the opportunity to extend business functionality by providing users, such as employees, suppliers and customers to access business logic through a browser, or commonly referred to as ‘business though a browser’.

E-Business applications and architecture uses a three-tier client server model, as illustrated below in figure 1. A three-tier client server model is where the client is mainly used for display with application logic and the business rules partitioned on the server, which is the second tier, and a database server is the third tier. (Chaffey, 2004).

![Figure 1: Three-tier architecture.](image-url)

Although e-Business can provide so many opportunities, many barriers may exist, preventing
an organisation adopting e-Business. The effective adoption and implementation of ICT may rely quite a lot on individual factors such as organisational size, structure, mix of available human and financial resources and capabilities (Pavic et al., 2007).

This is further supported by resource issues relating to finance, skills and personnel are expected to impact on the decision to adopt e-Business, (Fillis et al., 2004). Research has found that there are a number of barriers and success factors associated with e-Business adoption for an SME. (Taylor & Murphy, 2004) identified high initial set-up costs of ICT and e-Business can act as a barrier for an SME. Clearly, SMEs are presented with a number of challenges to realise the benefits e-Business that can provide to an organisation. However, open source technology may provide a solution by providing access to technology which may not be cost prohibitive.

1.1 Open Source

Open source is a development method for software that harnesses the power of distributed peer review and transparency of process. The promise of open source is better quality, higher reliability, more flexibility, lower cost, and an end to predatory vendor lock-in (Open Source Imitative OSI, 2008).

Open source software has grown in popularity in recent years, with governments, business firms, military organisations, and educational entities incorporating open source software into their enterprise functions to counter tightened budgets and rising operational expenses (Hedgebeth, 2007). Any reservations that exist regarding security may be short-sighted, with financial services institutions such as Experian, embracing open source software recently, (Ferguson, 2008). It appears open source popularity will continue to grow, with predications that by 2012, more than 90 percent of enterprises will use open source in direct or embedded form, (Judge, 2008).

Open source appears to be here to stay, but what open source software is available which can be used to develop e-Business applications, based on an e-Business three-tiered architecture (Figure 1).

2 TECHNOLOGY REVIEW

The research team identified various Database Management Systems (DBMS) which could be used in the data tier of the three-tier architecture. The research team analysed SQL Server, Microsoft Access and MySQL to consider suitability for e-Business applications for the case study.

Microsoft Access is a DBMS developed by Microsoft Corporation and belongs to the Microsoft Office family of products. Microsoft Access includes database functions that allow users to perform database queries and has some very good reporting interface. Microsoft Access is a flexible package and is suitable for organisations that have small database requirements and require few database records. In a database environment with multiple users, connections and queries, Microsoft Access will probably become unreliable and slow. However, Microsoft Access can be less expensive than some other DBMS on the market.

Microsoft SQL Server is another Microsoft package, but this product is primarily marketed towards large organisations with large database requirements. SQL server is an enterprise-level database system and is widely used in organisations that require a database that can handle high use, multiple transactions and offers high reliability. Microsoft SQL server is certainly a viable option for large organisations and the cost highlights this, while SQL server is an expensive option and requires a lot of training to be able to use the product effectively.

MySQL is an open source project which has been publicly available since 1996, but has a development history going back to 1979, (Welling & Thomson, 2003). MySQL is a relational database management system (RDBMS), which enables data storage, searching, sorting and retrieval. MySQL has many attractive features which include speed, ease of use, query language support, capability, connectivity, security and portability (DuBois, 2003). MySQL has become a popular database solution as its popularity within the database community has increased and would probably suit most database needs.

It should be noted that there are several database options available and the views expressed by authors, practitioners and application developers on available database systems varies significantly. Many developers and practitioners may have a bias towards database systems with which they may have had previous experience. It is clear that the choice of database system will depend on a number of factors such as price, speed, efficiency, etc. The database comparison table illustrated below in Table 1 highlights the differences in the most suitable database systems available.

The research team analysed suitable scripting languages for the business logic layer of the three-
tier architecture. The research team analysed ASP, ASP.NET and PHP to consider suitability for e-Business applications.

Table 1: Database Comparison.

<table>
<thead>
<tr>
<th></th>
<th>Access SQL Server</th>
<th>MySQL</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cost</td>
<td>Inexpensive</td>
<td>Expensive</td>
</tr>
<tr>
<td>Speed</td>
<td>Weak</td>
<td>Strong</td>
</tr>
<tr>
<td>Efficiency</td>
<td>Weak</td>
<td>Strong</td>
</tr>
<tr>
<td>Users</td>
<td>Few</td>
<td>Many</td>
</tr>
<tr>
<td>Connections</td>
<td>Few</td>
<td>Many</td>
</tr>
</tbody>
</table>

ASP is a server-side scripting language developed by Microsoft to run dynamically generated HTML pages. ASP was originally introduced with Microsoft’s server platform and utilises VBScript, a variation a Visual Basic Language, consequently both Microsoft products. Although scripting languages have been changing rapidly, many competing technologies have introduced new integrated features, while ASP has been lacking in this area. However, many business environments have Microsoft servers. This is probably a major driver for the continued use of ASP and ASP supports Microsoft SQL server, which is also widely used.

ASP.NET is part of Microsoft’s .NET platform and should not be confused with ASP. Although both are developed by Microsoft, the only other link between both technologies is that they both use VBScript. ASP.NET can be used to build dynamic web sites, web applications and XML web services. ASP.NET offers developers the ability to build pages using controls such as buttons or labels, with properties assigned to the controls. The controls produce segments of code, e.g. HTML for forms. This concept introduces the idea of all the code behind the page. This may reduce development time, but it is fair to conclude that programmers and/or developers will always need a deep understanding of the technology, even with the .NET concept of code behind the page.

PHP is an open source scripting language and was originally started as a hobby, this programming language has grown in popularity and use in recent years. It can be imbedded into HTML pages and allows communication with a database. A number of databases support PHP, which include MySQL, dBASE, Microsoft SQL Server, Oracle and many more. PHP supports most back-end databases, but it is most commonly known for its seamless integrations with MySQL. It can also run on a number of operating systems such as Windows, Linux, MAC OS and most varieties of Unix.

Furthermore, it is well supported by several PHP discussion groups and communities. PHP has grown from a set of tools for a personal home page development to the world’s most popular web programming language, and it now powers many of the Web’s most frequent destinations (Shiflett, 2006).

The choice of ASP, ASP.NET or PHP will ultimately come down to the application requirements and the hosting provider environment. The choice will also include the developers’ familiarity with programming language. If the developer is familiar with C, Java, JavaScript, etc, then PHP may be the best choice. If however, the developer has little experience of programming languages the choice may be the .NET concept of code behind the page. ASP.NET or PHP would be more suitable for e-Business applications, as software support and updates for ASP becomes redundant in the future. The technology comparison table illustrated below in Table 2 highlights the differences in the most suitable technologies available.

Table 2: Scripting Language (Source: Oracle).

<table>
<thead>
<tr>
<th></th>
<th>ASP.NET</th>
<th>PHP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cost</td>
<td>Free</td>
<td>Free</td>
</tr>
<tr>
<td>Platform Cost</td>
<td>Expensive</td>
<td>Free</td>
</tr>
<tr>
<td>Speed</td>
<td>Strong</td>
<td>Weak</td>
</tr>
<tr>
<td>Efficiency</td>
<td>Strong</td>
<td>Weak</td>
</tr>
<tr>
<td>Security</td>
<td>Strong</td>
<td>Strong</td>
</tr>
<tr>
<td>Platform</td>
<td>Strong</td>
<td>Weak (IIS Only)</td>
</tr>
</tbody>
</table>

A number of technology options exist for e-Business applications and only a handful have been analysed due to suitability and requirements for the e-Business applications to be developed in this paper. This paper investigates the use of open source software for e-Business applications as access to financial resources can be limited in an SME.

The research team found that MySQL would be the most suitable database option for each case study organisation due to cost, stability and security. MySQL is a true open source database management package. As a result, the research team decided that PHP would be the most suitable scripting language, as it works seamlessly with MySQL and, once more, it is an open source product.

Therefore, the research team agreed to develop the e-Business applications using MySQL and PHP for the data layer and business logic layer, respectively. MySQL or PHP do not require a commercial licence and all are battle-proven on high volume websites. Apache was selected to provide the web server environment. The apache web server
open source project was developed and as an open source HTTP server for operating systems such as Unix and Windows. Apache has been the most popular web server on the internet since 1996 (Apache, 2008). Apache is also the most popular web server and has the largest market share, used on over 50 percent of web servers, (Netcraft, 2008). The software selected for the development of e-Business applications are Apache, MySQL, and PHP are open source software and are known collectively as AMP.

3 CASE STUDY

Organisation A was a micro-size manufacturing organisation based in the UK. The research team also found that the case study company was experiencing sustainable growth, with a year-on-year sales change of more than forty percent. The case study found that the organisation was suffering from ‘growing pains’ in a number of areas of the business. The research team carried out interviews, process review meeting and structured walk-through to understand business operations and indentify improvement opportunities. It was found that the organisation used office applications and had a legacy system for stock management, but many of its key processes were paper-based and manual. It was found that the organisation was becoming inefficient in a number of its key processes. The research team and management reviewed all improvement opportunities to identify ‘urgent improvements’. A key process in the dispatch department was causing a bottle neck and affecting other areas of the business. This was identified for improvement actions. The process was know as the ‘verification process’, was critical to the business, and directly affects the organisations ability to service customers.

This process was a manual process. It required personnel in the dispatch department to validate the data on six documents against a master data sheet to ensure all the data on the documentation matches that on the master data sheet. Once complete, a manager would then be required to repeat this process to verify all the documentation was correct, and data was not overlooked accidently. This activity was required to be completed on every product. It was a critical step, as the data and information contained on the documentation comprised of product information, safety data and certified values.

Organisation B case study was a medium-size car finance broker based in the UK. This organisation has been established for some time, with a growing finance portfolio. The primary role of this organisation was to link car dealerships and finance organisations able to provide funding for vehicle leasing and purchasing, acting as an intermediary. The organisation used a proprietary off the shelf software package to support its operations. The software did allow some customisation, but it was found this had its limitations.

Through management briefings, conducting interviews and process reviews at each visit, it was possible to identify issues and problems with business processes. The research team identified issues with current business processes, data integrity, processing time and legacy system. The organisation rely on external parties, (car dealerships), to provide legible and complete information for finance proposals. It was found that proposals are hand written by car dealerships and then faxed to the contact centre for processing. Once received, employees were required to transfer the information on each proposal onto the legacy system. However, if an application has not been hand written in a legible format employees were expected to decipher information on the proposal, presenting the organisation with data integrity issues.

Furthermore, it was found that because proposals were faxed this affected the quality of information on a proposal. The organisation team were fully aware of this problem and were keen to resolve this issue through the use of technology. The research team identified this as an area of improvement, as improvements, or elimination of data integrity problems would provide the organisation with clear and accurate information and more efficient business processes.

The organisation offers products which require a decision of acceptance or decline within two hours. It was found that sometimes this is exceeded due to incomplete or incorrect data collection, which necessitated follow up actions and telephone calls by employees resulting in extended processing time. It was found that if the processing time extended beyond two hours then a dealership would source finance from a competitor, resulting in lost commission for the organisation.
4 RESEARCH AIMS & OBJECTIVES

The overall aim of this study was to determine ways in which open source software may be used to develop e-Business applications to improve business processes in each case study organisation. The research focuses on identifying suitable open source software to develop such applications. The research then moved onto proposing solutions utilising internet technology, and then developing e-Business applications. To achieve this aim, a number of key objectives would have to be met. These key objectives include:

- Identity open source software for e-Business applications.
- Determine ways in which solutions can be developed using internet technology.
- Develop e-Business applications for each case study organisation.
- Identify benefits e-Business provided each organisation.
- Use findings in the paper for the development of an e-Business framework for an SME adopting and implementing e-Business.

5 RESEARCH METHODOLOGY

The primary research for this paper is based on a case study of two UK based SMEs, the first is in the manufacturing industry, and the second is involved in the financial services industry. Two case study organisations were selected for this paper so that research could be completed on a manufacturing organisation and non-manufacturing organisation. The case study is action research lead in both cases. In first case study the research team spent more than 18 months working with the case study company on an e-Business development programme, to develop e-Business capabilities and implement operations management tools and techniques. An associate from the research team was based at the case study company throughout the programme.

The second case study company is a medium-size enterprise based in the UK. The organisation investigated the use of internet technology in both front and back end activities to improve business operations to minimise processing time and data integrity issues to improve competitiveness of the organisation.

The research team carried out several face-to-face interviews with stakeholders from case study A and B. The interviews with case study A were primarily carried out with key stakeholders, but this was then extended to include employees involved in business processes to better understand each process. Once each process was fully understood, the research team then began identifying areas for improvement through the development of a web-based system.

In respect of case study B, initial interviews were held with senior management (directors), and the Information Technology (IT) manager. The research team began with understanding existing business processes. This process enabled the research team to understand finance proposal process from initial enquiry to release of funds.

In case study A and B the research team was involved in the system analysis and development process for each web-based solution. In case study A, the research team designed web-based applications to improve existing process, data handling and reduce resource requirements. In case study B the research team designed a new e-Business application to link car dealerships and the case study company to improve operational activities and enable information sharing. In both cases the research team developed the e-Business applications using open source software, which had been designed for each case study organisation.

6 FINDINGS

Organisation A

It was decided that a web-based verification application would be developed by the research team to improve the ‘verification’ process for organisation A. The study had already identified that access to financial resources for technology development would be a constraint. This resulted in the stakeholders agreeing that the ‘quick win’ project should be developed through the utilisation of barcode and open source technology. This was the perfect opportunity for both the research team and the organisation to witness the viability and suitability of open source technology, which was a new concept to the case study company.

The web-based application would be developed so that barcode data on a product master data sheet could be scanned and stored on a database. The data required for validation would be stored in barcodes on product documentation when printed. An
employee could then use the verification application to complete data verification checks using a barcode scanner. The application would check the barcode data on the documentation, against the master data stored in the database, informing the employee of a pass, or fail. This would result in eliminating management from this process, as the employee would not be required to perform data verification, thus eliminating human error from the process.

The application successfully developed with open source technology (APACHE, PHP and MySQL). It was successfully implemented and replaced the existing manual process, thus improving the process significantly. It was found that the new application provided several benefits, such as more efficient business processes within the dispatch department because the current bottle neck had been eliminated. This in turn helped speed up the dispatch of goods from the department. Furthermore, through better utilisation of human resource management, more time can now be spent in other areas of the business, which could provide even more cost savings or help improve revenue streams through new product development.

The total project cost for the application was calculated at £4006.35. This included hardware, software, training and development costs. The new process provided a resource saving of fifty percent and a noticeable improvement in throughput time. This was calculated as a total cost saving of approximately £20K in the first year, providing an approximate five hundred percent Return on Investment (ROI) for the case study company.

It was found that the verification system provided a number of benefits which could not be fully quantified. Some of which included the ability of the case study to eliminate the bottleneck observed in the department. This in turn helped increase the throughput of goods in the dispatch department. Furthermore, through better utilisation of human resource management, time can be spent in other areas of the business, which could provide even more cost savings or help improve revenue streams through new product development. The process was de-skilled, resulting in any employee being able to perform verification using the web-based system, after initial training.

In addition, the development of the web-based application includes the storage of data, using the MySQL database. As a result, management can use the data to monitor the process and retrieve historical data to help make business decisions. Finally, the research team found that knowledge and understanding of e-Business application had been positively witnessed and experienced by the entire organisation. Management and personnel involved in the verification process have commented that without the adoption of the system, it would have been impossible to meet current customer demands using the existing approach.

Organisation B

The research team and employees from the organisation reviewed several web-based applications used by competitors and organisations providing financial services. This provided the initial concept of a web-based application to the organisation. The research team held focus group meetings to discuss, and identify desirable functionalities of the proposed web-based application. It was agreed that the application would also include finance quotation tools for dealerships to use whilst using the web-based proposal form. It was thought this would enhance value proposition to dealerships, thus encouraging adoption. The web-based proposal form would include validation functions to ensure that all required information was completed on the form before submission, thus ensuring data integrity and minimising errors, thus reducing processing time.

It was agreed to develop and implement the solution in phases. In the first phase the solution developed would store all the proposal information in a database. The proposal would then be encrypted and emailed to the organisation for processing, this would replace the fax. The next stage of development would look at identifying solutions to allow data sharing between different operating systems and programming platforms.

The organisation successfully implemented a web-based e-Business application using open source software (PHP & MySQL). The application improved information flows between the car dealerships and the organisation, replacing the paper passed proposal application process. The application was first released to field sales staff for review and testing, before being released to car dealerships. However, it was identified that the next stage of the development may be more difficult, as the application would be required to share data directly with legacy systems, posing a security risk. Furthermore, a dealership has requested information sharing between both organisations to simplify the proposal process. If the organisation can provide a solution, the dealership as agreed to complete all business transaction with the case study company. This has resulted in the proposal for the application to be developed further, into a web service.
7 DISCUSSION

Although both case study organisations had implemented e-Business applications successfully using open source software, the research team found that a number of barriers existed, which included access to skills, key personal and business culture issues.

It was found that Case Study A had little knowledge of e-Business, e-Business applications or the benefits and opportunities that could be realised by an organisation adopting e-Business. The research team managed to involve as many employees as possible in the development of the e-Business application, to help deepen knowledge and understanding and encourage buy-in at all levels of the organisation. It was also found that the organisation lacked the necessary skills and knowledge to develop e-Business applications. Moreover, the research team found that access to personnel required to implement and deploy the application was limited, due to personnel involved in other business activities. Once the e-Business application was successfully implemented, the research team arranged a feedback meeting to ascertain personnel experience of e-Business. It was found that the approach to involve personnel had been received positively, as personnel felt that e-Business knowledge and understanding was now evident throughout the organisation.

It was found that Case Study B Company did have knowledge of e-Business applications, but did not fully understand how internet technology could be deployed to improve operational activities. The research team found that the organisation did have an information technology department, which were highly skilled in a number of areas, but did not possess the necessary skills to develop e-Business applications. Moreover, it was found that access to personnel was limited, similar to the first case study organisation. Clearly, key stakeholders must appreciate that access to personnel is required, which may remove personnel from daily business activities. The project feedback meeting found that all involved had a positive experience and found open source could provide significant benefits to the organisation. This organisation is now looking to develop the e-Business application into a web-service for its dealership network. This will also be developed using open source software.

This paper had already indicated that access to software can be cost prohibitive, but this was overcome through the utilisation of open source software. However, a number of other barriers for each organisation were observed by the research team, and these need to be considered by academics and practitioners involved in e-Business development projects for SMEs. Although many barriers may exist, each organisation experienced e-Business benefits, as illustrated below in Table 3 and Table 4.

<table>
<thead>
<tr>
<th>Cost Saving</th>
<th>Process Effects</th>
<th>Culture Effects</th>
</tr>
</thead>
<tbody>
<tr>
<td>£20K (year 1)</td>
<td>Bottle neck removed</td>
<td>e-Business knowledge &amp; understanding</td>
</tr>
<tr>
<td></td>
<td>Data integrity</td>
<td>e-Business buy-in</td>
</tr>
<tr>
<td></td>
<td>Increased throughput</td>
<td>Positive experience of open source software</td>
</tr>
<tr>
<td></td>
<td>De-skilled process</td>
<td>Open-source buy-in</td>
</tr>
<tr>
<td></td>
<td>50% human resource saving</td>
<td></td>
</tr>
</tbody>
</table>

Table 3: Case Study A - e-Business Benefits.

<table>
<thead>
<tr>
<th>Cost Saving</th>
<th>Process Effects</th>
<th>Culture Effects</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Information sharing</td>
<td>e-Business buy-in</td>
</tr>
<tr>
<td></td>
<td>Legible data</td>
<td>Positive experience of open source software</td>
</tr>
<tr>
<td></td>
<td>Reduced processing time</td>
<td>Open-source buy-in</td>
</tr>
</tbody>
</table>

Table 4: Case Study B - e-Business Benefits.

8 CONCLUSIONS & FURTHER WORK

The case study found that SMEs face a number of challenges, and that e-Business can provide opportunities for organisations to improve business operations, information sharing and remain competitive. However, they may face several barriers to e-Business adoption, one of which is access to technology. It is found that access to new technology can be cost prohibitive, or existing legacy system costs may prevent further investment in technology. This may be overcome by the use of open source technology.

The research team found that open source software can be used for the development of e-Business applications, namely MySQL and PHP. E-Business applications were developed and implemented for two case study organisations, to
support business processes and enable information sharing. Supporting the view of (Bocij, 2003) e-Business is aimed at using information and communications technology to integrate and enhance processes. In a value-chain context, the efficiency of processes is enhanced, which should result in lower-cost, higher quality products so giving the customer better value.

Although open source technology may provide an SME with an opportunity to develop low cost e-Business applications, access to personnel with the skills, experience and knowledge of such technology is still required. Therefore, more research into access to open source skills is required to ensure that open source technology development can be supported in the future.

REFERENCES

AN ONTOLOGY-BASED ARCHITECTURE FOR MULTI-AGENT SYSTEM ENVIRONMENT

Roberto Paiano, Anna Lisa Guido and Enrico Pulimeno
Dipartimento Ingegneria dell’Innovazione, Salento University, Via per Arnesano, 73100 Lecce, Italy
Tel: +39 0832 297229, Fax: +39 0832 297279
{roberto.paiano, annalisa.guido, enrico.pulimeno}@unile.it

Keywords: Multi agent systems, ontology reasoning, domain independent, domain dependent.

Abstract: The increasing interest towards e-business systems, and thus the need of the companies to communicate with other companies in an efficient and flexible way, brings to a new way of thinking about information systems that open itself towards distributed systems where the exchange of information may happen. The problem is in the way that this exchange can be made both from the technology point of view and from the conceptual point of view. If from a side a recent technology for information exchange (DDS - Data Distribution Service) can help us from the other one it seem interesting the use of ontology as representation tool of application domain that must open itself to the communication. In this paper we present an architecture that is oriented to a multi-agent communication that uses ontologies and DDS as information exchange protocol. The architecture here presented must be repeat without difficulties on an e-business system assuming that each agent of the proposed architecture is a company in an e-business system.

1 INTRODUCTION

An e-business system is a system where involved companies communicate among them exchanging data in as much as possible efficient and flexible way with the goal to satisfy the final user. Two are the main problems that rise in this area:

- Conceptual problem related to the definition of a semantics of interchange of the data in order to assure the communication of the companies in heterogeneous environments;
- Technological problem related to the use of a suitable data interchange system useful to assure the communication between heterogeneous systems.

Both the problems are difficult to be faced and surely can be of great help the formal ontologies to solve the first problem (conceptual problem) and, relatively to the technological problem, it can come in help the DDS technology (Data Distribution Service) (DDS, 2001) standard OMG and surely useful not only for synchronous communication among companies but also for the asynchronous one.

Very often, in fact, the companies don’t need only information in real time: we can think, for example, to a business process that needs to have a data coming from another information system. The data requested from the business processes in order to go on with the execution are not immediately available. It is necessary a mechanism that allows to ask for the data and wait until the data is made available by some other information system that can provide it.

A base architecture useful to think about an e-business system based on a multi-agent has been presented in a research project lead by SSI (Space Software Italia) and the Department of Engineering Innovation of the University of Salento. The project was founded by Apulia region. Starting from a multi-agent system already made up from SSI in a demining system where each agent is represented by a robot, the goal of the project has been to enable the communication between agents through a level of intelligence made up by semantic web technologies, in order to define a run-time all the parameters to publish/subscribe depending on the operational context where agents work.

Naturally, when in the multi-agent system the parameters to publish/subscribe has been identify, it is necessary to avoid that they are manually published/subscribed in the DDS; it would be, therefore, useful that the classes (the parameters are published/subscribed in the DDS through classes) that materially define these parameters inside the DDS, are automatically produced.
The "intelligent" level, that it is important to add, it has to operate so that to allow "to understand", depending on the context, the parameters to publish/subscribe. To add this level, the formal ontologies are particularly useful. Not only the ontology but also the technologies of reasoning that allow from a side to define the context where the system operates and from the other to individualize the rules of action that each agent can take in order to answer to a well defined event.

The comparison with this kind of system and the e-business systems is immediate: every agent represents every information system involved in the information interchange. The "intelligent" level allows subsequently the different information systems to publish/subscribe in the DDS the data of interest to an appropriate system of reasoning that allows to identify the correct information.

In this paper we present a high-level architecture designed in the research project: the architecture is useful in order to allow the communication among the involved agents. As it will be clearer subsequently, the base idea will be that to make the application domain independent in comparison to the definition of the key elements that make possible the communication (DDS). Besides particularly interesting is the sharing of an only one knowledge base among several agents of the system.

After having introduced in the section 2 the state of the art to the multi-agent systems, in the section 3 will be introduced three alternatives considered as it regards the positioning of the knowledge base on the various agents that constitute the system. In the section 4 we present the architecture of the multi-agent system, finally, in the section 5, we present the conclusions of the paper.

## 2 BACKGROUND

The e-business systems have not been thought and developed, until now, according to a multi-agent logic. It is useful, however, to present the state of the art related to the multi-agent systems in order to understand the problem list related to the development of a system of this type.

In a multi-agent system we may speak about 4 main aspects:

- **decisinal aspects related to the agent:** what actions an agent undertakes depending on the external environment;
- **outsourcing of the execution:** possibility that more agents collaborate together, performing elementary actions to complete a complex task;
- **interactions among agents:** information interchange among agents in distributed environment;
- **Evolution toward component more and more endowed with autonomy:** agents that in full autonomy reach their own goals.

The multi-agent systems need a study of the nature of the interactions. It results therefore important the notions of collaboration and cooperation.

Interesting, in this paper, to underline that for the realization of a multi-agent architecture it is necessary to have a layer of communication and a layer of conceptual modelling.

- **Communication Layer:** currently the most qualified is standard seems to be FIPA-ACL (FIPA, 2002) (Agent Communication Language) created from the Foundation for Intelligent Physical Agents (FIPA). This standard is founded on the linguistic action theory, elaborated by John Searle (Searle, 1969). An important implementation of the FIPA standard is the framework JADE (http://jade.cselt.it) an open source platform for peer-to-peer agent based communication developed by Telecom Italia Lab.
- **Conceptual Model Layer:** it is very important to identify the domain where agents operate and it is important the dynamics of interactions between agents.

Currently, particularly interesting within the modelling of the multi agent systems it results:

- The **BDI Model** (Beliefs-Desires-Intentions) (Chang-Hyun, Guobin et. Al, 1969): it considers the agent environment belief, which is the result of its knowledge and perceptions, and a set of Desires. Intersecting these two sets, we obtain a new set of intentions, which can become actions.
- **Tropos** (Bresciani, Perini et. Al, 2004) a software development methodology founded on concepts used to model early requirements. In particular, the proposal adopts Eric Yu's modelling framework, which offers the notions of actor, goal and dependency, and uses these as a foundation to model early and late requirements, architectural and detailed design. The language used in Tropos for the conceptual modelling is formalized in a meta-model described with a set of UML class diagram.

In literature there are several papers that examine the software engineering paradigm applied to the multi-agents system, among them an article of Pratik K. Biswas (Patrik, Biswas, 2007) describes...
extensively the multi-agent system elements and their correlations, confronting them with UML paradigm that it reuses for their modelling.

Very interesting in a multi-agent system is the ontological approach very useful to provide an explicit and formal representation of the domain. This representation is simple to realize and it is simple to exchange between agents thanks the ontological languages such OWL (W3C, 2004).

3 USE OF KNOWLEDGE BASE IN MULTI-AGENT SYSTEM

An e-business system thought in the multi-agent terms it will have, for the complexity of the scenarios in which it will operate, an elevated degree of hardness and modularity. To reach this goal it is important to make more independent possible the various agents (nodes) of the system, that they need a continuous interchange of data in order to reach their own goal.

The first problem to face is that to add to a multi-agent system a layer of "intelligence" that is able to provide a good level of autonomy and facility of updating the system, besides it provides the possibility to describe the several events that are verified inside the system and to which the agents must answer. This level of intelligence is constituted by an ontology (opportune supported by a system of rules) that it results particularly useful to guarantee a complete understanding of the domain to all the interested agents, and accordingly one swifter change of his in case of changes or the insertion and management of unexpected events.

In the use of ontology in the e-business systems based in multi-agents it is had to analyze with attention as it must be defined and above all if this ontology has to be positioned only on a node inside the system or must be distributes on different nodes. We describe 3 hypotheses of work individualized defining among them the most suitable to the context in which we work.

3.1 Centralized Knowledge Base

The centralized system foresees that all the agents that cooperate make reference to a knowledge base centralized on only one agent. This system involves that, all the information push through the central node which stores them in the knowledge base and elaborates them.

The advantages of this approach are:
- the system is simple to manage;
- presence of a supervision node from which it is possible to access all the information and to provide precise commands to the other agents;
- The disadvantages are:
- The supervisor node results a critical Point of Failure, in fact if the supervisor node had to come less for some reason the knowledge base would result unreachable from the other nodes and, therefore, the overall system go down.
- Agents have little decision autonomy.

3.2 Knowledge Base Total Distribute

In a structure that introduces an high degree of distribution, all the agents work in independent way and the exchange of data it has the goal to make possible the coordination among the agents. All the agents are in communication exploiting a system of connection reliable (the DDS) that limit the Point of failure. At ontology level, for this scenario two different solutions can be identified, one that foresees to repeat the whole knowledge base of domain on every agent and the other is to foresees the distribution of the domain ontology so that to put on every agent only the part of knowledge base that interests the specific agent for the carrying out of his/her own role in the system.

3.2.1 Knowledge Base Replied on the Agents

In this structure, in which knowledge base is repeat on every agent every node of the system, knows the whole domain.

The advantages of this approach are:
- Non-existence of a Point of Failure;
- The autonomy degree of the agent enhance
- All the agents have peer decisional ability exploiting the knowledge base;

The disadvantages are:
- Every change on the knowledge base has to be repeat on every agent;
- The system needs a good structure of coordination;
- The complexity of each agent enhances.

3.2.2 Knowledge Base Distribute on the Agents

Another solution foresees the distribution on the several agents of the description of the domain that must be decomposed in modules depending on the specific criterions of competence. In other words, it deals with decomposing in several functionalities the ontology and to implement them on the several agents based on the demands of these.
The advantages of this approach are:
- Least redundancies of the data, every agent has only the information of which it has need;
- Absence of a single Point of Failure, and therefore greater independence of the robots;
- Possible changes to be brought to the ontological structure of the data base would be alone on the single agent and not on everybody.

The disadvantages are:
- Planning is complex because each agent is complex.
- Limited decisional ability for the agents.

3.3 The Selected Solution

The selected solution is to have a distributed system with a knowledge base repeat on every agent of the operational context. In this way the several nodes that operate in the system will be fully autonomous and able to develop his/her own task without depending in some way on the other participants to the mission, also communicating and exchanging data, in continuous way, with the other agents.

4 LOGICAL ARCHITECTURE

OVERVIEW

Before introducing the logical architecture conceived it is fundamental to clarify shortly the operation of the middleware of communication selected.

4.1 Use of the DDS as Tool for the Communication among Agents

The DDS is a useful tool proposed by the OMG that enable the communication within data-centric systems. The system of communication is asynchronous and it is based on the publishing/subscribing protocol: when agent that operates in the system has the necessity of a data, it makes a subscribing of it pointing out the data of which it has need; when some other agent of the system makes the data available, it effects a publishing of it. The agent that has previously effected the subscribing is able, to this point, to get the data of interest. The data publishing/subscribing is encapsulated in a class and send to the DDS. Since information may change depending on the particular operational context, also the relative publishing/subscribing class can vary depending on the particular data to send to the DDS.

4.2 Goal in the Realization of the Architecture

The idea that is at the base of the conceived architecture is born from two fundamental requisite:
- **Decouple** the decisional aspect related to the publishing/subscribing of well defined information, from the technological aspect tied up to the necessity to produce on the fly the useful classes to publishing/subscribing the parameters in the DDS
- To provide an high flexibility level to the system in this way to allow an **adaptation** of the same to the different operational scenarios where the architecture could operate, also without denying the consequential potentialities from the existence of a middleware able to guarantee the communication among the various agents.

Having decided to use a semantic base replied on every agent of the environment, we decided to decouple the component of the knowledge base of domain from that related to the management and representation of the concepts that describe the middleware of communication (DDS).

The proposed decoupling makes the system independent from the particular operational context, providing a flexible and easily adaptable structure.

The proposed architecture is in fig. 1.

**Figure 1: Logical architecture.**

First of all we observe the presence of two layers one called "domain independent" and one called "domain dependent". Within these two layers,
interconnected through a system of rules as will be detailed after, the presence of two knowledge bases is observed:

- **Knowledge Base Domain Dependent**: this module represents the KB specific of a special operational context.
- **Knowledge Base Domain Independent**: this module represents several elements that constitute the DDS.

We can observe within the layer "domain dependent" the presence of an "agent". It represents a generic agent that will contain both the knowledge base domain dependent and the knowledge base domain independent and that will make operation of publishing/subscribing of the information on the DDS.

### 4.3 Ontology Domain Dependent Module

The knowledge of the context, in which the agents work, and of its dynamics, it allows to be able to manage in the best way the events that can be verified in it. The module has the goal to represent the application domain in which the agents operate and the possible semantics relationships existing among the information that the various agents could exchange. In this module it will be present, therefore, this domain ontology that must have realized from the expert of domain that, better of everybody, it is able to define every aspect of it. In the knowledge base it is possible to find the specific information about the domain and their semantic relationships as an example if the domain is an information system that manage data about environment, the concept in the knowledge base will be the *sensor* used to obtain data or the *area* where data will be obtained and these two concepts will be related each other in order to provide the semantic link. For each sensor it will be possible to define the *period of time* when the data will be obtained and the threshold values defined in the specific context. These information will constitute a small part of the overall domain ontology that describe the specific information system.

### 4.4 Ontology Domain Independent Module

In this module there is the ontology that represents the specifications of the substratum of communication chosen for this architecture. This knowledge base reproduces, of fact, the OMG specification for the DDS: this knowledge base will be completed with the individuals, depending on the specific necessities, from the information obtained following inference, from the KB domain dependent. For example, if it is important, for the information system to publish/subscribe a well specific data obtained by the sensor, for example the *temperature* data, the information useful to make a topic for the temperature will be moved from the knowledge base domain dependent to the knowledge base domain dependent.

### 4.5 Jena Reasoner Module

The Jena reasoner module have the task to realize, through the application of ad-hoc SWRL rules written by the business expert, the deductions on the knowledge base that allow to intercept events that must happen and that they require, eventually, of a publishing/subscribing of information on the DDS with the goal to complete one determined activity. With an opportune analogy, every information system within an e-business system will deduce from the domain ontology what information to publish/subscribe in order to communicate with the other systems.

### 4.6 Rule Engine Module

The rule engine module will provide, through the opportune rules defined in the SPARQL language, to extract from the KB domain dependent the information to publish/subscribe and will add individuals to the ontological classes that represent, instead, the DDS (KB domain independent).

### 4.7 Java Class Generator Module

Since the system of publish/subscribe of the DDS founds him, of fact, on a system of publish/subscribe of classes, the presence of this module is fundamental because it able to produce, beginning from the individuals of the KB domain independent, the useful classes to make the operation of publish/subscribe. The class draws the information that the domain expert associates to every topics ("topics" are the elements that able the communication in the DDS) this information, through the rule engine module, is repeat in the KB domain dependent and from here used for the creation of the relative classes.

### 4.8 Description of the Information Flow

The information flow in the architecture can be so defined:
The agent captures the information stored in the knowledge base domain dependent; The Jena reasoner module makes the appropriate deduction through the SWRL rule engine; The RULE ENGINE module defines the mapping rules between the two knowledge base (domain dependent and domain independent). The java classes generator generates the .java file and the corresponding .class file; The .class file will be used to send information through the DDS.

It is important to observe that the separation between knowledge base domain dependent and knowledge base domain dependent increases the flexibility level of the system. Changing the context of business in which the system operates mean to define the knowledge base domain dependent and the system of reasoning (module rule engine and Jena reasoner) leaving the dynamics of communication made up through the DDS.

5 CONCLUSIONS AND FUTURE WORKS

In this paper we propose an architecture conceived for the communication of a data centric multi-agent systems that they use as middleware of communication the DDS recently proposed by OMG.

Particularly interesting is the analogy among the data-centric systems and the multi-agent systems applied to an e-business context. The analogy can be done thinking about every information system involved in the e-business environment as a single agent of the architecture here presented. In this way it will be possible to bring on a e-business system the whole efficiency and the flexibility that the proposed architecture introduces: the great advantage obtainable is the separation between the KB domain independent and the knowledge base domain dependent, another advantage is the possibility to produce in automatic the useful classes to make the data publishing/subscribing of the data.

Naturally, repeating the architecture proposed on an e-business system there are many aspects to clarify that brings to several research ideas.

The first problem is surely tied to the semantic interoperability: it is necessary to build a knowledge base that describes the whole domain in which every information system participates but as it is possible to effect the transfer of the information proper of every informative system (which those present in the database) in format compatible with those described in the ontology domain dependent? And what mean, practically, to describe a domain?

Parallels to these problems, the problem of reasoning is still open: few has been done in international scientific community in this sense so much that doesn't exist, until now, a standard language universally recognized for realizing the reasoning. In this architecture SWRL is used but this doesn't exclude, in a next future, the use of a language more efficient.

In every case, is interesting the idea at the base of the present paper that consists of using in an e-business system the formal ontologies and the mechanisms of reasoning.

ACKNOWLEDGEMENTS

We would thank SSI (Space Software Italia) company and Antonella Falzone for the tangible support.

REFERENCES

SECURE IT/TELCO ENVIRONMENT PLANNING MADE EASY
A Concept of a Tool for Planning Secure IT/Telco Infrastructure and Applications

Wolfgang Haidegger
SECUDE Global Consulting GmbH, Lassallestrasse 7b, 1021 Wien, Austria
Wolfgang.haidegger@secude-consulting.com


Abstract: This paper first motivates the necessity of a planning tool for IT infrastructure and applications, which allows the inclusion of security measures in an automated way. Then the author summarizes the requirements coming from legal and technical standards, which serve as a framework to assure the compliance of the results of planning activities with the respective applicable regulations. Next, a rough concept for realizing the planning tool is presented and finally conclusions are presented.

1 INTRODUCTION

The IT/Telco industry supports e-commerce with a selection of electronic business applications aimed at commercial transactions. The following list enumerates some of the better known business applications, but is by no means exclusive:

- electronic funds transfer,
- supply chain management,
- e-marketing,
- online marketing,
- online transaction processing,
- electronic data interchange (EDI),
- automated inventory management systems, and
- automated data collection systems.

From a technical point of view this paper distinguishes between the network services (also called “infrastructure”) and the application services, which are necessary to implement business applications as listed above. This distinction is important, as it mandates different technical and procedural security measures.

In addition to distinguishing the type of service necessary (infrastructure or application) to put together a business application, one is very often faced with a converged IT/Telco landscape concerning both types of services. This means one has to accommodate for the security needs of both circuit switched and packet oriented networks with their different transport and control planes, management protocols and application service philosophies.

The last two paragraphs show that security issues pertaining to information stored or transferred within networks for the purpose of doing electronic business can get arbitrarily complex both on a technical and a procedural level. At the same time the need for security grows as more and more personal (and sometimes very private) data is involved in e-commerce transactions. This is also reflected by the fact that compliance requirements to national and international regulations concerning confidentiality, availability and integrity of information become stricter.

Taking the complex surroundings just described into account, then in order to carry out a proper analysis of security requirements and planning of the according security measures a comprehensive tool (or tool chain) needs to be developed, which guarantees

- technical correctness,
- compliance to all relevant regulations and
- proper tailoring to the business needs

for the solution found.

A final remark: The author of this paper does not distinguish between Greenfield analysis and analysis of existing infrastructure and applications concerning the capabilities of the planning tool, as this does not seem to be relevant for a preliminary treatment of the topic.
2 THE NORMATIVE FRAMEWORK

The following paragraphs show the steps necessary for risk analysis and mitigation. These have been described in different levels of detail but with the same general intention in ISO/IEC FDIS 27001, ISO/IEC 15408 and ASIS.

Figure 1 gives an overview over the procedure the steps are embedded in.

Step 1: Risk Analysis
- Assets, their value and their owners are identified.
- Vulnerabilities of the assets as well as threat agents, which might exploit the vulnerabilities, are identified.
- For the resulting risks possible countermeasures are identified and the appropriate ones are selected.

Step 2: Definition of Mitigation Measures
- Development of security policies
- Analysis of relevant international, national and corporate legal and technical standards relevant for the situation.
- Development of procedures and guidelines, designing the way the security policies shall actually be realized.

Step 3: Integration into Service/Infrastructure architecture (not shown in figure 1)
- Design of infrastructure with selected security measures
- Design of applications according to selected security measures

Step 4: Verification (not shown in figure 1)
- Selection of appropriate Assurance Level
- Development of Evaluation Assurance Level Criteria corresponding to the Assurance Level selected

The following subsections give an idea about the legal and technical recommendations the tool (or tool-set) will have to be able to abide to. They represent additional, project external constraints.

2.1 International Legal, Procedural and Technical Standards

There are many international legal standards, which have to be considered as important when treating IT/Telco security. Some examples are:
- Treaties of the European Union,
- European Convention for the Protection of Human Rights and Fundamental Freedoms,
- European Directives

To show one important European standard in the area of Identity Management:

The European Directive 95/46/CE: deals with data protection, is aimed at giving to the data subject (owner of data) the most control possible on its own identity and personal data, posing a series of requirements on recipients, controllers, processors and even third parties. Art. 2, letter a), giving a definition of "personal data", says: "identifiable person is one who can be identified, directly or indirectly, in particular by reference to an identification number or to one or more factors specific to his physical, physiological, mental, economic, cultural or social identity".

There are also quite some procedural (non-technical) international standards, which will play a role when planning security measures. Two examples are:
- ISO/IEC FDIS 27001: This International Standard specifies the requirements for establishing, implementing, operating, monitoring, reviewing, maintaining and improving an IT security management plan focusing on the overall business risks of the organization the plan is made for.
- COBIT: The Control Objectives for Information and related Technology (COBIT) provides a set of generally accepted measures, indicators, processes and best practices to assist
  - in maximizing the benefits derived through the use of information technology and
  - in developing appropriate IT governance and control in a company
International technical standards will come from ISO, IETF, ITU-T, OMA, TMF, W3C and other international standardization bodies or industry forums.

2.2 National Legal Standards

Every nation will have a set of national legal standards, which have to be considered as important in addition when treating IT/Telco security. This time the examples are selected from the USA, again with the focus on relevance to Identity Management:

- Privacy Act of 1974: all government agencies - federal, state and local - which request social security numbers are required to provide a disclosure statement on the form;
- Family Educational Rights and Privacy Act (FERPA, also known as the "Buckley Amendment," enacted in 1974, 20 USC 1232g): social security numbers fall within the scope of personally identifiable information that is restricted from disclosure by schools that receive federal funding under the Family Educational Rights and Privacy Act;
- Children's Online Privacy Protection Act (COPPA) - 15 U.S. Code 6501 et seq.: The act's goal is to place parents in control over what information is collected from their children online;
- Financial Services Modernization Act, Gramm-Leach-Bliley (GLB), Privacy Rule - 15 USC 6801-6827: The 1999 federal law permits the consolidation of financial services companies and requires financial institutions to issue privacy notices to their customers, giving them the opportunity to opt-out of some sharing of personally identifiable financial information with outside companies;

2.3 Corporate Standards

Finally, organizations whose security needs to be planned or evaluated and updated will probably have additional regulations, which need to be taken into account. For example it might be necessary for every employee to wear a badge with her or his photo id on it to guarantee an additional possibility to identify her or him.

3 CONCEPT FOR THE REALIZATION OF THE PLANNING TOOL

The threat analysis will be based on a database that holds information about known threats to the specific system components, their architecture, protocols. This database will also be filled with the results of the tasks dealing with risk assessment. There are tools today, which do this on a very general level (e.g. COBRA). The main challenge will be to drill down to implementation level if possible and/or necessary.

Furthermore a network/application planning tool will be developed based on the threat database above, which is capable of identifying threats to a planned network, suggesting methods to mitigate the threats according to a specific Evaluation Assurance Level and including the results into the network plan, a threat model document and a test specification. Here two issues will be of main concern:

- Developing solution variants, which really fit the topic and providing guidance for the selection of the "right" solution.
- Using the selected Evaluation Assurance Level as the driving parameter for the automated solution development

The tool will be capable of tracking the changes to the planned network, the according changes in the threat model, and the changes in the security measures to be taken and finally, track the changes in the documentation (network plan, test specification).

Figure 2: General architecture of the tool (tool chain).

Figure 2 give a first impression of how the tool chain shall be designed. The main part will be an intelligent central data base, which will serve all planning applications likewise.
4 CONCLUSIONS

The paper shows the necessity of a general IT/Telco security planning tool (or tool chain) as a consequence of heightened complexity concerning the infrastructure and the applications as well as tightened security and compliance regulations.

The regulatory framework, which such a tool chain has to accommodate, is dealt with conclusively concerning the types of regulations, but only exemplary concerning the actual recommendations.

Finally a first idea of a possible realization is given together with the crucial points to solve.

Next steps will be the collection of the technical ingredients of the database (protocols, HW architectures, SW architectures, SW frameworks, ...) and a first design of the database itself. This will have to happen in the light of the applications, which will use the database.

REFERENCES

WEB AND MOBILE BUSINESS
SYSTEMS AND SERVICES
FULL PAPERS
MOBILE TOURISM SERVICES
Experiences from Three Services on Trial

Niklas Eriksson and Peter Strandvik
Institute for Advanced Management Systems Research, Åbo Akademi University, Joukahainenkatu 3-5A, Turku, Finland
nikeriks@abo.fi, peter.strandvik@nim.ax

Keywords: Mobile tourism services, technology adoption.

Abstract: For this study a field trial was conducted to identify the determinants for tourists’ intentions to use three trial services targeting tourists on tour, in this case on the Åland Islands in Finland. We identified that the major barrier for the non usage of the trial services was linked to the type of travel that the trial group participated in. Also price transparency and ease of use especially ease to take new mobile services into use should be highlighted in mobile tourism service development. Moreover, we came across some basic reminders to take into account when commercializing mobile services, such as carefully define a customer target group, estimate potential usage volume and plan marketing / sales tactics. These aspects are not necessarily realized enough in technology development.

1 INTRODUCTION
The use of the Internet for doing commerce or interacting with customers has been growing rapidly in the world wide tourism industry. Mobile commerce, or e-commerce over mobile devices, on the other hand has had many conflicting predictions on its future popularity. Most predictions have been overly optimistic. However, the benefits that arise from mobile technology have not yet been fully delivered, which to some extent is explained by the fact that mobile applications, due to complexity or lack of relevance, fail to meet customers’ expectations (Carlsson et al. 2006). Travel and tourism is an industry in which several different projects have been conducted where mobile applications have been developed, tested and implemented, some even with moderate success (e.g. Ardissono et al 2003, Kramer et al 2005, Schmidt-Belz et al 2003, Repo et al 2006). Some of these pilot projects (e.g. Kramer et al 2005, Schmidt-Belz et al 2003) have been focusing on GPS which the average tourist doesn’t yet have in his/her handheld mobile device. Therefore it seems relevant to build and test services that actually can be used by the average tourists. Nevertheless previous pilots have given us valuable information on the potential of mobile technology.

The New Interactive Media (NIM) project, with funding from the European Union and the regional government of the Åland islands, is a development programme of increasing knowledge, production and use of new interactive media on the Åland Islands in Finland. Within the project several mobile applications have been developed for the travel and tourism sector on the islands. Three of these services will be presented more in detail in this paper: MobiPortal, TraveLog and MobiTour. A field trial of these services with real incoming tourists to the Åland Islands using their own mobile phones has also been conducted. Findings and experiences from this trial will be reported. Possible determinants for consumers’ intentions to use mobile tourism services will be discussed as well.

2 SERVICE DESCRIPTIONS
The services have been planned with a common logic namely the Braudel rule: freedom becomes value by expanding the limits of the possible in the structures of everyday life (as presented by Keen & Mackintosh 2001). The rule is then translated into a tourism setting which means that tourists’ real or perceived need has to be met by the services and

1 Åland is an autonomous and unilingual Swedish region in Finland with its own flag and approximately 26.700 inhabitants. Åland is situated between Finland and Sweden and consists of 6 500 islands. (www.visitaland.com)
moreover, the services need to profoundly change the way a tourist does or experience something – and to the better (Harkke 2007).

MobiPortal is a mobile version of an information portal www.visitialand.com which is the official tourist site of the Åland Islands. The portal includes search for events, restaurants etc., a map service and facts on the Åland Islands.

TravelLog is a mobile community for incoming tourists to share experiences from the Åland Islands with each other. The virtual meeting place includes stories, pictures, tips and interactions.

MobiTour is a guide for attractions such as the Bomarsund fortress which is downloadable / streamable to the visitors’ own devices. The guide includes voice and/or video guidance.

All these three services ought to expand the limits of a tourist to the Åland Islands according to the Braudel rule by enabling 1) instant access to local information, 2) enhanced communications with other people with the same interests and 3) experience enhancement for certain unmanned attractions. Especially experience enhancement features are generally seen as key drivers for successful customer satisfaction in tourism (Pine & Gilmore 1999). The determinants for consumer usage of mobile tourism services are, however, a complex issue which will be discussed next.

### 3 POSSIBLE DETERMINANTS

Several models of technology adoption have been developed. One of the most used models is the technology acceptance model (TAM) by Davis (1989) which is based on the theory of reason action (TRA) by Fishbein et al. (1975). Other often used models in technology adoption research are the diffusion of innovations theories (DIT) by Rogers (1995) and the unified theory for the acceptance and use of technology (UTAUT) by Venkatech et al. (2003) which combines TAM with other acceptance model e.g. DIT. Here different components of these models will be discussed, together with relevant research theories for adoption of electronic and mobile services, to identify possible determinants for consumer intentions to use mobile tourism services.

The TAM model proposes two determinants, perceived usefulness and perceived ease of use, which impact the acceptance of technology and adoption behavior as a result (Davis 1989). Perceived usefulness is defined as “the degree to which a person believes that using a particular system would enhance his or her performance”.

Perceived ease of use is defined as “the degree to which a person believes that using a particular system would be free of effort”. The two TAM determinants are proposed to identify the intended usage behavior of a system and are widely used as a backbone for research in adoption of technology. However, the first TAM variable perceived usefulness is foremost designed to research work performance improvements in organizational contexts. In consumer markets consumer behavior is also influenced by other factors. It is typical that non-efficiency factors impact consumer adoption of technology, e.g. good tourist technologies are not only those that make tourists more efficient, but that also make tourism more enjoyable. Thus tourism can be characterized as wandering, where tourists attempt to enjoy the city environment and chance upon things of interest, rather than optimizing (Brown & Chalmers 2003). As the mobility (on the move) capability is generally seen as the key value driver in m-commerce (Anckar & Eriksson 2003), mobile technology clearly has the potential to support the wandering aspect of tourism. A word like flexibility has commonly been used to describe the independence of time and space that is provided by mobile technology. According to Kim et al. (2005) the hedonic motivation or the enjoyment aspect of tourism has, however, not been clearly defined in mobile technology acceptance models. The perceived type and degree of perceived value of a mobile service depend on the other hand on the situation or context of usage (Mallat et al 2006, Lee & Jun, 2005). Anckar & Dincau (2002) introduced an analytical framework that identifies the potential value creating features of mobile commerce. Mobile value elements in the framework for consumers on the move are: Time-critical arrangements, Spontaneous needs, Entertainment needs, Efficiency ambitions and Mobile situations. Time-critical arrangements refer to applications for situations where immediacy is desirable (arise from external events), e.g. receive alerts of a changed transport schedule while on tour. Spontaneous needs are internally awakened and not a result of external events, e.g. find a suitable restaurant while wandering around. Entertainment needs, killing time/having fun, especially in situations when not being able to access wired entertainment appliances, e.g. kill or fill time in transportation. Efficiency ambitions aim at productivity, e.g. use dead spots during a travel to optimize time usage. Mobile situations refer to applications that in essence are of value only through a mobile medium (e.g. localization services), which ought to be the core of...
mobile commerce. Consequently *perceived mobile value* represent the degree to which a person perceives value arising from the mobility of the mobile medium.

Nevertheless not only the medium creates value for the consumer but the essence of the services as well. We refer to such value as *perceived service value*. For example for a tourist in a planning or booking situation the key to successful satisfaction would be timely and accurate information relevant to the consumer’s needs (Buhalis 2003). Equally important for a tourist visiting a historical attraction may be the satisfaction of educational and entertainment (edutainment) needs (HyunJeong & Schlusser 2007). Similarly a person with a mission to share experiences with others may find satisfaction when a community responds (Arguello et al 2006). The three examples refer to the essence of the three services on trial.

The second TAM determinant *perceived ease of use* has been widely discussed in mobile commerce. Limitations of mobile devices (e.g. screen size) cause consumers to hesitate whether to adopt mobile commerce or not. According to Cho et al (2007) device limitations suggest that focusing on easy to use mobile applications could enhance the consumer acceptance of mobile commerce. Kaasinen (2005) points out that mobile services need to be easy to take into use as well as mobile services are typically used occasionally and some services may be available only locally in certain usage environments. As a consequence, information on available services should be easy to get and the services should be easy to install and to start using. The ease of taking a service into use may in fact have a direct impact on the adoption behaviour of a mobile service (Kaasinen 2005). On the other hand when problems arise, users in the consumer market are often expected to solve the problems on their own (Repo et. al 2006). Consequently the use may rely on proper instructions or on a helping hand from someone. Proper support conditions also in a consumer market may therefore be important especially for advanced mobile services. Nevertheless consumers many times expect to take a new product or service into use without instructions or help.

According to Rogers (1995),"The innovation-decision is made through a cost benefit analysis where the major obstacle is uncertainty". *Perceived risk* is commonly thought of as felt uncertainty regarding possible negative consequences of using a product or service and has been added to the two TAM determinants as a negative influencer on intended adoption behaviour (Featherman & Pavlou 2003). Trust, as trust in the service vendor to minimize the risks, has also been added to the TAM model (e.g. Cho et al 2007, Kaasinen 2005) and pointed out as a strong influencer on the intended use of mobile services due to that mobile commerce is still at its initial stage (Cho et al. 2007). We refer to trust as the perceived risk defined by Featherman & Pavlou 2003. They divide the perceived risk for electronic services into the following elements; performance risk, financial risk, time risk, psychological risk, social risk and privacy risk. Performance risk refers to the possibility of a service to malfunction and not performing as it was designed and advertised. The financial risk refers to the potential monetary outlay associated with the initial purchase price as well as the subsequent maintenance cost of the product and the possibility of fraud. Time risk refers to that the consumer may lose time when making a bad purchasing decision e.g. by learning how to use a product or service only to have to replace it if it does not perform to expectations. Psychological risk refers to the potential loss of self-esteem (ego loss) from the frustration of not achieving a buying goal. Social risk refers to potential loss of status in one’s social group as a result of adopting a product or service, looking foolish or untrendy. Privacy risk refers to the potential loss of control over personal information, such as when information about you is used without your knowledge or permission. At least security and privacy issues have been highlighted as barriers to mobile commerce (O’Donnell et al. 2007). Also financial risks in form of high costs, including operating costs and initial costs, have been highly ranked by consumers as hindrances for m-commerce in its early stages (Anckar et al. 2003).

In UTAUT *social influence* among other constructs is added to the two TAM components and defined as the degree to which an individual perceives that important others believe he should use the new system (Venkatech et al., 2003). Social influence is also known as subjective norm in the theory of reason action (Fishbein et al 1975) and in its extension theory of planned behavior (Arjzen 1991). In consumer markets image and social status have been proposed to impact consumers’ adoption of mobile services (Teo & Pok 2003). Also the number of users may influence, especially for community services which usefulness heavily depend on activity of different participants (Pedersen et al 2006). Furthermore other external sources such as media reports and expert opinions
may influence consumers’ perception of electronic services (Bhattacherjee 2000).

Demographic variables such as gender and age are commonly used in consumer research. For example, gender and age might through other constructs influence the intended adoption behavior of mobile services (Nysveed et al. 2005). According to the theory of planned behavior (Ajzen 1991), control beliefs constitute individuals’ belief that they have the necessary resources and knowledge to use an innovation. For example skills or earlier experience of using mobile services may influence the adoption intentions of new mobile services. When discussing consumer behavior in tourism and the impact of information and communication technologies (ICTs) a clear distinction should also be made between experienced and inexperienced travelers (travel experience). The first group mainly feels more comfortable organizing their holidays and thereby taking advantage of ICT tools available to them more easily (Buhais 2003). Moreover, inexperienced destination travelers usually need a lot more local information. Innovations also need to comply with the existing values and needs of the individual in an everyday life setting (Moore & Benbasat 1991), in this case while on tour. For example the values of the individual may differ depending on the type of travel they are on: leisure or business, where the former ought to call for services with enjoyment rather than efficiency. In consumer markets mobile services also compete against existing and constantly developed alternatives. Thus consumer habits are usually quite slow to change from known alternatives (Dahlberg & Öörni 2007). People are on average risk-averse. But that is not true for everyone as we have individuals who are earlier to adopt new ideas than others (Rogers 1995). Such personal characteristics make diffusion of innovations possible. Personal innovativeness is the willingness of an individual to try out and embrace new technology based services. Individuals’ limited mobile device readiness has as well been seen as a great negative influence of the usage of more advanced mobile services (Carlsson et al. 2004). We refer demographic variables, experience of mobile services, travel experience, destination experience, type of travel, personal innovativeness and user device readiness as discussed here to tourist characteristics as they illustrate key characteristics of an individual that may influence the intended use of mobile tourism services.

Based on the literature discussion possible determinants for consumer intentions to use mobile tourism services are: perceived mobile value and service value, perceived ease of use, social influence, perceived risk and tourist characteristics. Mobile value and service value replace perceived usefulness as presented in the TAM model. Ease of use is defined as in the TAM model where also ease of taking a service into use is included. Social value is defined as in UTAUT and perceived risk as presented by Featherman & Pavlou (2003). Tourist characteristics constitute key characteristics of an individual on-tour. The defined determinants are summarized in Table 1.

Table 1: Possible determinants for consumer intentions to use mobile tourism services.

<table>
<thead>
<tr>
<th>Determinant</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mobile value</td>
<td>the degree to which a person perceives value arising from the mobility of the medium</td>
</tr>
<tr>
<td>Service value</td>
<td>the degree to which a person perceives value arising from the essence of the service.</td>
</tr>
<tr>
<td>Ease of use</td>
<td>the degree to which a person believes that using a particular service would be free of effort</td>
</tr>
<tr>
<td>Risk</td>
<td>the degree to which a person feels uncertainty regarding possible negative consequences of using a service.</td>
</tr>
<tr>
<td>Social influence</td>
<td>the degree to which an individual perceives that important others believe he should use the service.</td>
</tr>
<tr>
<td>Tourist characteristics</td>
<td>Demographics, Experience of mobile services, Travel experience, Destination experience, Type of travel, Personal Innovativeness, Device readiness</td>
</tr>
</tbody>
</table>

4 THE FIELD TRIAL SETUP

According to Repo et al. (2006) TAM theories and similar approaches have little relevance in the real product development process. Product developers need first hand user feedback in form of personal interaction rather than by reading research reports. The arguments are based on experiences from piloting a mobile blog service for tourists, where the user gave direct feedback to the developers orally and through survey forms. Involving the consumer in the development process of products or services can be very rewarding indeed (von Hippel 2005). With the theoretical foundation (Table 1) in mind and with the idea of directly interacting with the consumers to receive direct and spontaneous feedback to the product developers we designed a field trial which included oral, observed and survey data collection.

The trial was conducted during a conference in the capital of the Åland Islands Mariehamn 21 –
22.9.2007 at the legislative assembly where the main activities of the conference were held. The conference was arranged by the local Junior Chamber of Commerce organization and it was called WestCongress2007. Members of similar organizations in the western regions of Finland were invited to attend the conference. A total of 191 participants had registered in advance for the conference. The trial was coordinated in cooperation with the conference director who offered assistance with e.g. stand preparations and informing the participants in advance of the mobile services in conference guides, online and during registration.

Our stand was set up at the main entrance of the building where the main activities were held. The main entrance was the place that we anticipated would be the busiest during the first parts of the conference when we were invited to promote and demonstrate our services. The stand was equipped with a video projector showing animated picks of the services and also flyers, tables and chairs for comfortable discussions with the conference attendants.

At our stand the conference participants were informed more in detail of the services. The services were also demonstrated, which gave us a chance to observe peoples first time reactions. The stand also provided for us a good place to freely discuss different issues regarding the services with the participants. Participants filled out voluntarily a questionnaire which also was an agreement to contact them by e-mail after the conference to follow up on their own independent use of the mobile services during their stay on the Åland Islands. Each phone and operator connection (device readiness) was checked by the stand representatives to ensure that the participants actually were able to use their own phones for the services.

In the questionnaire the participants were asked to fill out questions according to the constructs defined for tourist characteristics:

- **Demographics**: Gender and age
- **Experiences of mobile services**: Commonly used services were listed with the alternatives: 
  - [1] continuously using 
  - [2] have tried 
  - [3] have never tried.
- **Travel experience**: How often they travel for more than one day:
  - [1] several times a month
  - [2] ~ once a month
  - [3] 3 – 9 times a year
  - [4] < three times a year.
- **Destination experience**: If they have visited the Åland Islands before:
  - [1] Yes, > 5 times
  - [2] Yes, 2 – 5 times
  - [3] Yes, once
  - [4] Never and their knowledge of the Åland Islands


- **Type of travel**: if they consider WestCongress2007 to be:
  - [1] a leisure trip
  - [2] a business trip
- **Personal Innovativeness**: Three statements were proposed on a five point scale:
  - [5] definitely agree - [1] definitely disagree: I want to get local information through my mobile phone when...
    - 1. I plan my program e.g. in the hotel.
    - 2. I’m on my way to a local place with e.g. bus.
    - 3. I get acquainted with a local place on foot.

Kristoffersen & Ljungberg (2000) distinguish between three types of mobility: visiting, traveling and wandering. Visiting, an actor performs activities at different locations (e.g. a hotel). Traveling, an actor performs activities while moving between different locations usually inside a vehicle (e.g. bus). Wandering, an actor performs activities while moving between different locations where the locations are locally defined within a building or local area (e.g. on foot).

For the follow up a semi-open web questionnaire was used to receive feedback on the participant’s actual use of the three services. The web questionnaire was sent to the participants by e-mail two days after the conference finished ensuring that their service experience would be fresh in their minds. A reminder was sent a week later. The participants were asked to state for each of the three services whether they had used it or not. Their answer was followed up with an open question on their primary motivation for using or not using the service. In the analysis the answers were interpreted according to the theoretical foundation on determinants for the intended use of mobile tourism services. Additionally the participants were asked to state what kinds of problems they had run into if problems occurred. The participants were also to state on a five point likert scale ([5] Yes, definitely - [1] Definitely not) for each service what their intentions are to use similar services in the future while visiting a destination. Finally the participants were free to comment on the service.

5 **THE PARTICIPANTS**

Members signed up in advance for the conference were 191 in total. However, about thirty persons
didn’t register. We estimated that about 50 persons visited our stand. Out of these 50 persons voluntarily and without a prize draw 23 filled out the questionnaire and an operator connection (device readiness) that allowed them to use the services. Thereby it was relevant to send the follow up by e-mail to these 20 persons. Two mail addresses did not respond. Out of the 18 persons that the follow up went to 9 answered it.

Of the 23 that filled out the questionnaire 12 were men and 11 women. The average age was 35. The majority (66%) stated that they travel about once a month for more than one day. Most of them (66%) had visited the Åland Islands before at least two times or more. However, a majority (66%) answered that they know the Åland Islands satisfactory or not at all. Almost all (96%) felt the WestCongress2007 to be a leisure trip. Of the 23 participants all had at least at some point tried to use a mobile service and a clear majority (66%) used at least one mobile service continuously. A total of 74% (17) of the participants answered that they want to get local information with their mobile phone for at least one of the statements in table 2.

6 EXPERIENCES FROM THE TRIAL

To draw peoples’ attention to our stand we really needed to sell the services. As people were moving for different things in the building and to other locations in the surrounding area a major job was to get them to stop by the stand. Very few participants stopped without a few sales lines from the stand representatives, although they were informed in advance of the services and the stand was strategically placed at the main entrance.

Most people who visited the stand expressed a positive response by the first sight of the services. Comments like “that seems practical” and “I already use mobile news services so why not use these services” were given. Especially MobiPortal awakened concrete interest as it was bookmarked by a couple of stand visitors. A few persons also praised the visual design of the MobiTour guide. However, some people were spontaneously skeptic about the long download times for MobiTour. Nor did anyone ask for transactions over Bluetooth although it would have been possible at the stand. Several persons instantly also asked for the price of the services. The services were not charged for and it seemed like the transaction costs were obvious to most visitors and not a hindrance to use, except for the large files of MobiTour. Connection problems occurred with at least one network operator which interestingly led to that a few thought there was something wrong with the trial services.

None of the nine respondents to the follow up had on their own used any of the trial services. All reported that their primary motivation for the no use was that they didn’t experience a need to use the services during their stay at the conference on the Åland Islands.

The future intended use of similar services as the ones on trial were reported as shown in table 3. Services similar as MobiPortal received the highest score.

7 DISCUSSION

A customer target group needs to be defined for each mobile service developed (Hoegg & Stanoevska-Slabeva 2005). The primary target group for the three mobile services on trial is visitors to the Åland Islands. The trial targeted participants of WestCongress2007 who visited the Åland Islands. When analyzing the trial group it can be said that it was both right and wrong. It ought to be the right group based on the fact that most participants who filled out the questionnaire had a device readiness (87%) that allowed the services to be used on their own phone. The group already continuously used mobile services to a great extent (66%) and thereby the barrier to take on new services ought to be lower. Their knowledge of the Åland islands was only

Table 2: Local information with mobile phone.

<table>
<thead>
<tr>
<th>I want to get local information through my mobile phone when...</th>
<th>N</th>
<th>%*</th>
<th>Tot 23</th>
</tr>
</thead>
<tbody>
<tr>
<td>I plan my program e.g. in the hotel</td>
<td>10</td>
<td>43</td>
<td></td>
</tr>
<tr>
<td>I’m on my way to a local place with e.g. bus, car</td>
<td>14</td>
<td>61</td>
<td></td>
</tr>
<tr>
<td>I get acquainted with a local place on foot</td>
<td>11</td>
<td>48</td>
<td></td>
</tr>
</tbody>
</table>

* 5 = definitely agree and 4 = partly agree

Table 3: Intended use of similar services.

<table>
<thead>
<tr>
<th>When visiting a destination in the future I intend to use...</th>
<th>N</th>
<th>Mean*</th>
</tr>
</thead>
<tbody>
<tr>
<td>Similar services as MobiPortal</td>
<td>9</td>
<td>3,33</td>
</tr>
<tr>
<td>Similar services as TravelLog</td>
<td>9</td>
<td>2,89</td>
</tr>
<tr>
<td>Similar services as MobiTour</td>
<td>9</td>
<td>2,89</td>
</tr>
</tbody>
</table>

satisfactory or none (66%) which ought to create a need for local information. Also their willingness to get local information in different situations (74%) with their mobile phone was positive. Moreover the group was an experienced group of tourists (66%) which generally is found to be positive regarding usage of information and communication technology. On the other hand the group had a ready made program during the weekend and we observed that they also asked their hosts for tips and directions. The need for local information and guidance may therefore have been satisfied. Moreover they had their conference group who they met with continuously to share their experiences with. Consequently the service value of the three services on trial was already met by other means of interaction.

The analyses of the trial group indicate that the same people but with another mission to visit the Åland Islands could be a potential user group of the services on trial. The mobile value of using mobile services is, as discussed in the theoretical foundation, very much situation based. Moreover, the proposed value needs to comply with the user’s existing on-tour values. In this case self arrangement values by using a mobile phone necessarily didn’t exist due to the packaged set up of the conference. Consequently the type of travel, as packaged or non-packaged, is therefore to be taken into account as an influencer of the intended use of mobile tourism services. A non-packaged tour ought to comply better with an individual’s values of self arrangement / service. Nevertheless customized mobile services aimed at specific needs of packaged groups such as conference attendants may indeed generate value.

The trial also shows that we cannot forget that new technology innovations very seldom sell themselves. Much of our efforts at our stand were sales related. Launching new mobile services certainly need to be pushed by creating awareness among the potential consumers as for any other new product. Similar pointers have been presented by Collan et al. (2006): “Hot technology doesn’t sell itself, it has to be marketed to the consumer in the shape of value adding services that are easy to use”. Therefore marketing / sales tactics influence needs to be set as a determinant for consumer intentions to use mobile tourism services.

Questions on the prices of the trial services were the most frequent ones asked during the trial. Therefore it seems that the financial risk is carefully accounted for by the consumers in their intentions to use a mobile tourism service. In this trial the services were free of charge and the transaction costs didn’t seem to be a barrier. Nevertheless our experience from this trial is that the service price and potential transaction costs must be transparent to the consumers to minimize uncertainty of the monetary layout. The monetary aspect may be even more important for foreign visitors as transaction cost may rise noticeably.

Even though many participants expressed a general interest in the services it is also a fact that no one reported that they actually used the services on trial. Thereby questions are raised from a business point of view on the potential usage volumes of the services on trial at this time and place. We certainly need to be very realistic when we launch mobile services on the potential volume of usage, especially when setting the business logic (Collan et al. 2006). Moreover product developers need to remember to look at things from a consumer perspective. For example in this trial the consumers thought the trial services didn’t work because of an operator problem. In the eyes of the consumer this means a malfunctioning product which is useless. Similarly long download times to access a service for a temporary use may cause the consumer to view the service as too time consuming to take into use. Neither can we expect consumers to install services in advance as according to Kaasinen (2005), “users are not willing to spend their time on something that they do not get immediate benefit from.” Consequently the ease of use aspect must be highlighted by product developers as mobile tourism services may be only temporarily used during a visit to a destination or a local place.

8 CONCLUSIONS

This paper presented possible determinants for consumer intentions to use mobile tourism services. A major driver according to the six identified determinants; mobile value, service value, ease of use, risk and social influence and tourist characteristics couldn’t be determined as no one used the trial services on their own. The major barrier for the non usage of the three services among the trial group seemed to be linked to the value aspect of the packaged tour (type of travel). Based on the experience from this trial we propose that researchers and practitioners especially take the following into account:

- The type of travel is a key aspect in designing mobile tourism services
• Marketing / sales tactics influence should be highlighted as a determinant for consumer intentions to use mobile tourism services
• Price transparency is an important aspect to minimize consumers’ perceived risk of mobile tourism services
• Ease of use aspects should be highlighted even more for mobile tourism services as they may be only temporarily used

These pointers can also be seen as reminders in technology development where basic commercialization routines sometimes aren’t realized enough. As for any other product defining a customer target group, estimate potential usage volumes and plan marketing are vital steps in launching mobile tourism services.

It needs, however, to be kept in mind that the experience is based on only one field trial and therefore further research in evaluating mobile tourism services and similar mobile services is needed. The recruitment of trial users could as well be done differently. According to Kaasinen (2005) ideally users should be allowed to use the trial services freely but it may lead to, as in this trial, to a no usage. Therefore some rules on minimum trial times should be set up, where additional usage to the minimum can be considered as real usage. Logs can also be helpful in data collection to receive prompt service usage data in addition to follow up data from the respondent. Moreover, phone interviews may give more extensive answers and better response rates in a follow up data collection of the same character as in this trial.

REFERENCES

Information Technology & Tourism. vol. 9, no. 3-4
Cognizant, USA.
ADOPTION VERSUS USE DIFFUSION

Predicting User Acceptance of Mobile TV in Flanders

Tom Evens, Lieven De Marez
Research Group for Media & ICT (MICT-IBBT), Ghent University, Korte Meer 7-9-11, Gent, Belgium
Tom.Evens@UGent.be, Lieven.Demarez@UGent.be

Dimitri Schuurman
Research Group for Media & ICT (MICT-IBBT), Ghent University, Korte Meer 7-9-11, Gent, Belgium
Dimitri.Schuurman@UGent.be

Keywords: User research, adoption diffusion, use diffusion, mutual shaping, mobile TV.

Abstract: In the contemporary changing ICT environment, an increasing number of services and devices are being developed and brought to end-user market. Unfortunately, this environment is also characterized by an increasing number of failing innovations; confronting scholars, policy makers as well as industry with an explicit need for more accurate user research. Such research must result in more accurate predictions and forecasts of an innovation’s potential, as a basis for more efficient business planning and strategy implementation. However, the success of a new technology is not only depending on the adoption decision and the number of people actually buying it, but relies at least as much on its actual usage. Hence, the focus of truly user-oriented acceptance or potential prediction should focus on predicting both adoption diffusion and use diffusion. Within this paper, we illustrate the added value of such an interactionist approach for the study of future adoption and usage of mobile TV by the assessment of both a large-scale intention survey and qualitative techniques such as diary studies, focus group interviews, observational and ethnographic methods.

1 INTRODUCTION

In the contemporary changing ICT environment, an increasing number of services and devices are being developed and brought to end-user market. Unfortunately, this environment is also characterized by an increasing number of failing innovations; confronting scholars, policy makers as well as industry with an explicit need for more accurate user research. Such research must result in more accurate predictions and forecasts of an innovation’s potential, as a basis for more efficient business planning and strategy implementation.

In most cases however, this need for more accurate user insight only gets translated in a cross-sectional investigation of the innovation’s adoption potential. However, the success of a new technology or service is not only depending on the adoption decision and the number of people actually buying it. For example, many people may have bought or adopted a mobile phone with GPRS, UMTS or MMS without using the feature. The success of an innovation is thus not only depending on its adoption, but at least as much on its usage. Hence, the focus of truly user-oriented acceptance or potential prediction should not only be focussed on predicting adoption diffusion, but also on predicting use diffusion and potential usage. Evidently, the first research question to answer remains up to which degree the innovation has the potential to be adopted. This should always be accompanied with an answer to the question up to which degree the innovation also has the potential to acquire a place in people’s and household’s daily lives (in terms of time and habits).

In terms of theoretical frameworks, the first ‘adoption diffusion’ question relies on the diffusion paradigm, while the second ‘use diffusion’ question relies on the ‘social shaping’ and ‘domestication’ paradigm. Too often however, the Social Shaping of Technologies (SST) and Domestication perspective is considered as the alternative to set off the lack of attention for the user and his/her social usage context in the diffusion theory. Traditionally, both perspectives (and the research based on them) have too much been considered as opposites; while they
are perfectly complementary to each other. The purpose of this paper is to illustrate this complementariness and the enrichment of combining the more quantitative generalizing research approach of diffusionism with the more qualitative in-depth SST research approach. Based on user research conducted on mobile TV, we illustrate how this combination of approaches and methods resulted in a prediction of potential as well as usage of this new technology. This way, we intend to illustrate the theoretical, methodological, managerial as well as policy relevance of this plea for a more mutual shaping or interactionist approach on predicting user acceptance (see Boczkowski, 2004: 255).

2 TWO COMPLEMENTARY FRAMEWORKS

The oldest of the two theoretical frameworks is the ‘diffusion framework’, of which Everett M. Rogers (1962) is assumed to be the founding father. According to this framework, the diffusion of innovations in a social system always follows a bell-shaped normal distribution, in which there can be successively distinguished between Innovators (2.5%), Early Adopters (13.5%), Early Majority (34%), Late Majority (34%) and Laggards (16%). A person’s innovativeness is assumed to be determined by the perception of the following set of innovation characteristics: relative advantage, complexity, compatibility, trialability and observability (Rogers, 2003). Since the early 60’s the theory’s assumptions on segment sizes, diffusion pattern and determinants have been a basis for different types of (mostly) quantitative research such as econometric diffusion modelling or innovation scales (Goldsmith & Hofacker, 1991; Meade & Islam, 2006; Moore & Benbasat, 1991; Parasuraman & Colby, 2001; Venkatsh, Morris, Davis & Davis, 2003).

Since the mid 80’s however, questions about its technological determinism and lack of attention to the user and usage of the innovation have induced Rogers to adjust his approach to the adoption decision process, but have also led to the rise of new paradigms such as domestication focussing on the ‘way the use in households is being socially negotiated and becomes meaningful, within the social context of class, gender, culture or lifestyle’ (Van Den Broeck, Pierson, Pauwels, 2004: 103; Haddon, 2007; Silverstone & Haddon, 1996) or ‘the process of taming and house training ‘wild’ technological objects, by adapting them to the routines and rituals of the household and thus giving them a more or less natural and taken-for-granted place within the microsocial context of that household’ (Frissen, 2000: 67; Jankowski & Van Selm, 2001: 37). Domestication thus refers to integration of new technologies in the daily patterns, structures and values of users, relying on a more social determinism (Bouwman, Van Dijk, Van den Hooff & van den Wijngaart, 2002).

Methodologically, the SST and domestication paradigm relies more on a qualitative tradition of methods such as in-depth interviews, ethnographic observation and diary studies.

In the past, these two major paradigms have mostly been regarded as opposite and competing, with convinced advocates from the two sides engaging in vicious debates. However, with diffusionism as the more quantitative tradition with the focus on acceptance and adoption decisions and the domestication tradition as more qualitative with a focus on the use and appropriation of technologies, both paradigms are clearly complementary (Punie, 2000). Or, as Boczkowski (2004: 255) states, ‘two sides of the same innovation coin’. To date a dialectical approach, which considers the development and diffusion of ICT innovations as ‘joint processes of technological construction and societal adoption’ (Boczkowski, 2004: 257), gains ground. Instead of thinking in terms of diffusionism or social shaping, the mutual shaping or interactionism approach (Boczkowski, 2004; Lievrouw & Livingstone, 2006; Trott, 2003) appeared in the late 90’s as a dynamic middle path between the two previous linear deterministic predecessors. By integrating both quantitative and qualitative research outcomes within this paper, we aim to illustrate the enrichment of such an interactionist approach for the development and rollout of mobile TV in Flanders, the northern and Dutch-speaking part of Belgium.

Relying on the difference between ‘adoption diffusion’ and ‘use diffusion’ (Shih & Venkatesh, 2004), we believe that the prediction of ‘adoption diffusion’ should rely on (1) a quantitative diffusion approach by means of (intention) surveys and modelling to gain insight in the innovation’s potential in terms of percentage of the target market, penetration pattern and profiles of the different adopter segments; and (2) the prediction of ‘use diffusion’, based on more qualitative techniques such as diary studies, focus group interviews, observational and ethnographic techniques (if possible in a field trial or living lab setting).
3 RESEARCH DESIGN

The empirical findings are based on the two-year MADUF project which studied the possibilities of mobile TV using DVB-H in Flanders. In first instance, a large-scale user survey (n: 575) was set up in order to forecast the market potential, or to predict the ‘adoption diffusion’ potential for mobile TV in Flanders. By applying the Product Specific Adoption Potential (PSAP) scale, we were able to map the size and nature of the future mobile TV market in Flanders. The PSAP scale is an intention based survey method in which respondents are allocated to Innovator, Early Adopter, Majority and Laggard segments based on the stated intentions on a general intention question and on respondent-specific formulated questions gauging for their intention for ‘optimal’ and ‘suboptimal’ product offerings (De Marez & Verleye, 2004; Verleye & De Marez, 2005). The scale was compared on its reliability with five other adoption models and has been applied to and validated for a diversity of ICT innovations such as digital TV, 3G, mobile TV and mobile news (De Marez, 2006; De Marez, Vyncke, Berte, Schuurman & De Moor, 2008).

In second instance, a representative panel of test users was randomly selected from the 575 survey respondents to experiment with mobile television devices in a ‘living lab’ setting during two weeks. Due to practical reasons (the DVB-H network was operational in the city of Ghent only, so the panel contained people exclusively living but not especially working in Ghent) and because of the rather explorative nature of this field trial, the amount of test users was limited to 30. With this field trial, we aimed at achieving a first realistic view of how future users will integrate mobile TV in their everyday practices. Users were asked to document their experiences in diaries while logging their activities, noting their comments and taking pictures of their usage situations.

Next to these data, we also gained insight in their personal evaluation of the trial phase by means of a post-measurement. Comparing these results with the findings of the market forecast before testing the device allowed us to see whether user attitudes towards mobile television had changed as a result of the trial. In this manner, we aimed to measure the effect of trialability, the degree to which an innovation may be experimented with on a limited basis (Rogers, 2003: 266). Explanations for possible shifts between the pre- and the post-measurements can be found in the usage diaries and two organised focus groups. Figure 1 illustrates this interactionist approach combining both quantitative user attitude research and qualitative ethnographic techniques.

4 RESULTS: PREDICTING ADOPTION DIFFUSION

By applying the PSAP scale to 575 rich cases, we obtained a reliable view on the size and nature of the various adoption segments for mobile TV in Flanders in the following segmentation forecast. While traditional fixed segment sized methods are reflected by the black line (in this case Rogers’ Diffusions of Innovations), the red line represents the adoption potential for mobile TV. The latter is contrasted to the potential of 3G (De Marez, 2006), which allows TV programmes to be received over a unicast architecture network. Figure 2 clearly shows that there is little demand for mobile TV over DVB-H compared to Rogers’ full market approach and even compared to the take-up of 3G services. Due to the lack of substantial innovative segments (Innovators and Early Adopters), we would recommend a partial market approach or even a niche strategy for the introduction of mobile TV in Flanders. This implies a specific introduction strategy for a limited market potential to serve the chosen segments in an optimal manner (about a 20% market penetration). Since the Late Majority and Laggard segment are clearly not willing to pay for this mobile service, we will define the maximal target group as Innovators, Early Adopters and Early Majority promising a 16,7% segmentation forecast.

In general, we witnessed a rather dual profile within the innovative segments with on the one hand well-earning, older executives (little time, potential...
for snacking) and on the other hand low educated young couples without children (much time, complementary to heavy TV viewing behaviour). Although especially executives are facing a shortage of time, most of them seem to be heavy television viewers, watching both entertainment and information programs. Especially Innovators and Early Adopters (joint for statistical reasons) possess advanced mobile phones (with camera, MMS, WAP, MP3, FM radio…), which they use in an innovative manner (e.g. sending e-mails on mobile phone, see Figure 3). Generally, these people show the highest willingness to pay for mobile TV while most of them consider a mobile TV device (with integrated mobile phone) as a substitute for their current mobile phone.

Figure 3: Sending e-mails on mobile phone

Clearly, such quantitative research may provide reliable estimations of the adoption potential and diffusion (in this case of mobile TV in Flanders), but does not provide us with in-depth information regarding the domestication and potential use diffusion of mobile TV. What place will it take in the lives of the consumers, how and when will it be used?

5 RESULTS: PREDICTING USE DIFFUSION

To answer the latter questions, one needs a more qualitative ‘use diffusion’ and domestication oriented research framework. In the case of mobile television a combination of diaries, focus group discussions, pre-test comparisons and photo elicitation within the boundaries of a living lab setting was used to get further insight in people’s usage patterns of mobile TV. Although we are aware these results are not statistically representative due to the very limited sample of 30 test users, they nevertheless allow us to identify some explorative usage patterns for mobile TV amongst our field trial participants.

On average, people watched approximately eleven times via their mobile television device during the two-week test period. However, it is possible that people being part of a panel within a test environment felt obliged to experiment more with the devices than they would do within a more natural context. Although we cannot ignore this trial effect, it plays a less important role within this research set-up because we aim to generate explorative rather than statistically representative findings. In terms of this usage frequency pattern, we can distinguish three kinds of viewers: light viewers watching less than 10 times (n: 15), medium viewers watching between 10 and 20 times (n: 13) and heavy viewers watching more than 20 times (n: 2). These two heavy viewers were identified as Innovator and Early Adopter within our large-scale sample.

Within our user panel, we only found two heavy viewers while the rest of the panel was about equally divided among medium and light viewers. One important finding during our test period is that the different types of viewers used the mobile TV device in a different way. Figure 4 represents all watching moments and divides them amongst the periods people watched mobile TV. In terms of the moments people watched mobile TV, we identified six different time slots: night (0-6h), morning (6-12h), noon (12-14h), afternoon (14-18h), evening (18-22h) and late evening (22-24h). When analysing the figure, we see that, except for the light viewers, trial participants are not inclined to watch mobile TV while having breakfast. This is probably due to the strong position in the morning of the medium radio, which is ‘together with the water and the stove, the first thing that is turned on in the morning’ (Winocur, 2005: 325). Light viewers are also more likely to watch mobile television at noon while having dinner.

Figure 4: Usage patterns (per time slot).

Heavy viewers are most likely to watch mobile during the afternoon, while most of the other types of viewers only switch their device on in the evening after coming home from work or school (see Figure 4). While light and medium viewers are watching mobile TV in the evening, we notice a remarkable
decline in viewing of the heavy viewer-segment during this time slot (see red line). Nevertheless, we see that this segment starts watching again in the late evening, the moment where the other segments switch their device off. This results in peaked watching patterns that differ quite much between the three user segments. While light and medium users show one viewing peak during the evening, heavy viewers have two peaks: one in the afternoon and one in the late evening. The latter two-peaked pattern is rather complementary with traditional TV, as its peak time comes right in between the mobile peak times. We can conclude that heavy viewers used mobile TV complementary to their regular television and therefore watched the device in a manner it was meant to be watched: on the move. In contrast, light and medium viewers watched mobile TV at home as a substitute for regular television.

The previous findings are supported by the usage locations indicated in the diaries. Light and medium viewers especially watched mobile TV at home. Undoubtedly, the most popular place was the living room where people are used to watch regular television while relaxing in their sofa. This also seemed the case for mobile TV: most people watched television in their natural habitat. Instead of watching the large screen, our test users watched mobile TV, albeit for a rather short period. After having tested the mobile device, they switched to the large screen again to enjoy their favourite programs. Here, we witnessed a substitution of the classical screen at traditional peak times with mobile TV was considered a second TV (see also Schuurman et al., 2008). This was especially the case for the light and medium users in our sample. This does explain the similarities between peak times for mobile and regular TV for these groups.

Another popular location for watching mobile TV was the kitchen. People seem to enjoy watching mobile TV while eating in the kitchen, where most of the time no TV set is at hand. We also witnessed that a lot of people used the mobile device while working at their desk or sitting behind the computer. These people used mobile TV rather as a background medium or as tertiary activity (see Jacobs, Lievens, Vangenck, Vanhengel & Pierson, 2008). When they heard something interesting, they switched attention from their work to the mobile device. Although they watched mobile television, these people considered the mobile television device often as a radio, which is in most cases also used as a background medium. Here mobile TV was clearly used in combination with other activities such as doing the dishes or working (multitasking).

Especially heavy viewers made use of the complementary function of mobile TV and considered it as an extra supply next to their regular television. This is illustrated by the fact that heavy viewers watched significantly more in public space and on the move. We found that watching in the car is a rather popular activity to kill time, sometimes as fellow passenger but also as driver. These people driving to their work and back, spend a lot of time in their car and have to suffer traffic jams. It is hardly surprising that in such cases mobile television is seen as a simple time killer although the radio can serve this purpose as well. Other persons preferred watching mobile TV while waiting for or travelling with public transport services (bus, metro and train). Taking into account the massive success of the iPod, mobile TV devices can be the next big thing to spend time while commuting.

After the trial period, we asked our 30 test users to fill out the same questionnaire they had previously taken. Based on the combined results of both pre- and post-trial measurement, we were able to compare the findings and see whether user expectations and attitudes had changed during the mobile TV field trial. The findings from the qualitative part of this research project (i.e. focus groups and ethnographic methods such as usage diaries) enabled us to explain possible shifts.

General interest for mobile TV slightly increased during the field trial. However, persons who originally intended to purchase a mobile TV device soon, now preferred to wait a bit longer. On the other hand, the amount of people certainly not willing to purchase a mobile TV device declined as well. A slightly increased average score (from 3,70 to 3,80) suggests that overall attitude towards mobile TV became a little bit more positive. Also the average price people are willing to pay increased from €233 to €294. But it is striking that we witness a converging shift towards a non-decisive average. Convinced believers start to doubt while disbelievers might have seen some possibilities after all due to the trial.

In other words, less people are showing an innovative attitude towards mobile TV, but many others shifted from ‘never’ to ‘maybe’. It thus seems that the field trial has raised awareness of mobile TV and that a lot of people do not consider the medium as a luxurious product any longer, making it less appealing to the more innovative but more likely to consider for the less innovative. Although these people are not likely to purchase mobile TV soon, they are not longer against mobile TV since they have experienced it as a handy medium to catch up television content quickly. Innovators and early adopters on the other hand were somehow disappointed by the lack of interactive and
interesting content, resulting in their downgrade. Despite the shift towards a more positive attitude, the potential for mobile television remains dramatically low, as the sample does not contain any Innovators or Early Adopters anymore and that the least innovative segments (Late Majority and Laggards) remain largely overrepresented.

6 CONCLUSIONS

With this paper, we intended to reconcile two opposing traditions: adoption diffusion and use diffusion. Within the MADUF-project, we combined research techniques from both traditions in an interactionist way in order to get a more holistic view on the possible success of mobile TV in Flanders. By means of a PSAP-estimation, it became clear that mobile TV is not ready yet for total market acceptance so that a partial market or even niche strategy was suggested. By means of a diary study, combined with a pre-test and post-test survey during a mobile TV-trial in a living lab environment, we were able to get a better understanding of the possible use diffusion of mobile TV. We found that for most test persons traditional television remains the reference point for evaluating mobile TV. Television undoubtedly is one of the most domesticated technologies within the home and became so dominant that people often schedule their behaviour in function of the TV-set. We found that light and medium mobile viewers used the device at home as a second TV with watching behaviour in line with traditional TV. Heavy users on the contrary watched mobile TV in a truly mobile and much more complementary way with traditional television. This resulted in mobile peak times coinciding with regular TV for the former two groups, while for the latter mobile TV allowed to extend the regular TV viewing peak with two mobile peaks: one before and after the regular peak. Finally, we witnessed the (modest) overall positive effect of trialability through a slightly increased general attitude towards mobile TV during the field trial.

By combining these two paradigms, we were able to draw a clearer picture of the potential success of mobile TV and the different factors influencing this success. While a quantitative potential estimation can identify adoption segments and describe them for targeting purposes, the qualitative usage diffusion-research provides input for the fine-tuning of the technology in terms of usage patterns, features and content. We believe this methodological plea for more interactionist research designs has theoretical as well as industry and policy relevance for the prediction of ICT user acceptance. For instance, in the current debate of digital dividend such predictions could help policymakers to get insight in the feasibility of new communication technologies and for which new technologies they should preserve space in the future radio spectrum. These estimations also allow marketing managers to decide in which market segments they should invest and with what offer these segments should be targeted. Finally, for researchers we hope this paper gives some food for thought about the added value of an interactionist approach and inspires them to work out more creative innovations research designs in the future.

ACKNOWLEDGEMENTS

The MADUF-project (Maximize DVB Usage in Flanders) was supported by grants from the research centre IBBT (Interdisciplinary Institute for Broadband Technology) and a consortium of both broadcasters and network solution companies. The MADUF project aimed to maximize the social and economic valorisation of DVB-H for the Flemish citizen, government and industry (broadcasters, operators and constructors) through the development of a technological and regulatory consensus model (pax mobilis).

REFERENCES


Keywords: Adoption, Mobile service, TAM (Technology Acceptance Model), TPB (Theory of Planned Behaviour), J2ME, Finland.

Abstract: This paper offers both theoretical and empirical contributions relating to the consumers’ motives for and barriers to adopting mobile services. A case study of a mobile service is introduced to validate the factors that facilitate and trigger, and respectively hinder the acceptance of mobile services in Finland. The factors affecting the service adoption were researched based on a conceptual model that was a modified version of the Technology Acceptance Model (TAM) and of the Theory of Planned Behaviour (TPB). Based on the findings from a survey conducted with a sample of 52 Finnish campus students, the factors that affect consumers’ attitudes towards mobile services were discovered. Usefulness was revealed to have the strongest impact on attitude towards a mobile service, with context information and attitude acting as significant determinants of the intention to use the service. This study attempts to help mobile service providers choose right development and marketing strategies for mobile services and enhance the adoption and diffusion of mobile services.

1 INTRODUCTION

For a long time Finland has been a European pioneer for testing new mobile services. But even in Finland, the adoption of mobile services has been modest and the most popular services are still based on SMS messaging. Current mobile services do not fulfill the requirements set by the users since the services have been developed based on technological development instead of users’ needs. According to Moore and McKenna (1999) the mobile services are at the chasm, the pioneering and early adopters are using the services, but the masses have not yet followed their lead. Many researches (Hyvönen and Repo, 2005; Bouwman et al., 2007) have studied the factors that affect the adoption of mobile services in the Finnish context but there is still a need to evaluate the adoption factors in real-life case studies. Kaasinen (2005) has studied the user acceptance of mobile services by using the Technology Acceptance Model (Davis, 1989) and built user acceptance on four factors: perceived value of the service, perceived ease of use, trust, and perceived ease of adoption. Bouwman et al. (2007) have researched mobile service adoption from the perspectives of physical, cognitive, security and economic barriers, and from the perspectives of the perceived entertainment value and perceived flexibility benefits of mobile services. Finnish consumers’ usage of mobile services has been studied in a report for National Consumer Research Centre by Hyvönen and Repo (2005). The report focuses on mobile services that Finnish people use, the motives for and obstacles to using the services, and the readiness to pay for mobile services.

The success of mobile services is not only depending on technological maturity of mobile technology and the penetration of mobile handsets; in the end, eventually the users decide on the success of the services. The predictions are hard to make, but there are certain factors that need to be considered when developing mobile services. Muller-Veerse (2000) has identified ubiquity, reachability, security, convenience and personalisation as the key drivers for the mobile market today. Rogers (1995) has developed ‘Rogers’ basic five’ factors, later increased with two more factors (Moore and Benbasat, 1991) that influence the adoption and diffusion of new technologies throughout a social system: relative advantage, compatibility, complexity, trialability, observability, image and
trust. According to Rogers mobile services are likely to be adopted if mobile solution has some clear advantages over existing products or services. Mobile services should also be compatible with existing practices, values, and skills of potential adopters, and easy to understand and use. If it is not possible to experiment with the service before making an adoption or rejection decision, and if the benefits offered by the service are not visible, then adopters are likely to perceive uncertainty and risk in adopting the service. Today the mobile phone is more like a reflection of one’s identity, so if the adoption and use of mobile service is perceived to communicate one’s personality and enhance the image and status, it has more potential to be adopted. It is also important that the mobile service adopter perceives the service provider to be trustworthy.

There are several factors inhibiting the use of services, although the technology maturity is no longer an obstacle and even so the adoption and diffusion of mobile services has not yet proceeded as fast and wide as have been assumed. Development of technology is commonly seen as an important prerequisite, or even as a reason, for people to adopt new mobile services. According to Gillick and Vanderhoof (2000), the new mobility experience, i.e. the anytime and anywhere access to desired products and services, will be the greatest benefit for the consumer. However, Figge (2004) claims that the ubiquity alone is not enough to ensure a user acceptance and the success of a mobile application: as the user’s problems vary with the context in which he or she accesses service, the service based on information of the specific context is likely to become the user’s favourite. Gilbert and Han (2005) identified one key barrier of attracting a critical mass of adopters as the lack of compelling content, and another as the differences among adoption patterns in individual market segments. Aarnio et al. (2002) proposed two major reasons above all inhibiting the use of mobile services; the high pricing and the restricted quality of services.

Hyvönen and Repo (2005) have stated that mobile services are not adopted linearly according to the assumptions of diffusion theory. Their studies state that the socio-demographic factors do not essentially affect the adoption of mobile services; instead, the best way to explain the use of mobile services is to research the general attitudes towards new technology, mobile services and devices. In a survey by Cheong and Park (2005) conducted in Korea, perceived playfulness was found to be influential in predicting the behavioural intention to use mobile Internet. Whereas findings from a consumer survey conducted in Finland (Anckar, 2002) indicated that m-commerce adoption mainly appears to be driven by a need for solutions that add convenience and flexibility to daily routines rather than excitement and entertainment. Anckar (2002) also found out that consumers perceive the ability to satisfy spontaneous and time-critical needs as the most important driver of m-commerce adoption.

This paper researches and validates factors that facilitate and trigger, and respectively hinder, the adoption of mobile services in Finland. Finland has been selected as a case country since it is considered as a pioneer of mobile technologies. In August 2007, 97% of Finns had at least one mobile phone (Statistics Finland, 2007). The findings from the survey conducted in autumn 2007 (Invest in Finland, 2008), based on information on four million mobile phones from three telecom operators, showed that Finns change their mobile phones every 2.7 years. More than 70% of the phones in use had a colour screen and half included a camera. The 3G phones gained an 18% share of the market, which was more than double the figure in 2006. In the study in which participated over 1500 Finns aged 15-64 in year 2006, one quarter of those surveyed reported having browsed web pages with their mobile phones (It-viikko, 2007).

Furthermore the paper studies what kinds of services are likely to be adopted; and how businesses can enhance the adoption and diffusion of mobile services. In order to validate the adoption factors in practice this paper presents a real life case study of launching the mobile service. The case mobile service was based on downloadable Java application, not yet eagerly adopted at least in Finland. It was tested in a delimited user group of students and personnel in Finnish campus area in the field trial. Mobile service offered information about internal news, events, announcements and schedules of the study community, as well as a mobile flea market for buying and selling own things through mobile channel. The aim of provided services was to enable easy access for essential information despite the time and place. Especially students require multi-channel services due to their mobile lifestyle.

This paper is organized as follows: chapter 2 represents the conceptual model of mobile service adoption and chapter 3 describes the empirical evaluation of case mobile service. In chapter 4 the feedback questionnaire of the case study is presented and the results of the survey are analysed. The paper concludes with the limitations of the survey and the suggestions for future work.
2 CONCEPTUAL MODEL OF MOBILE SERVICE ADOPTION

In our study, we developed a modified version of the Technology Acceptance Model (TAM) (Davis, 1989) and the Theory of Planned Behaviour (TPB) (Ajzen, 1991) for researching the adoption of mobile services. According to TAM, a consumer’s behavioural intention to use a “system” is determined primarily by its usefulness and ease of use. In TPB, the intention to use a “system” is explained by attitudes toward a certain behaviour, subjective norm and perceived behavioural control. TAM and TPB have both been widely used among researchers and found to be very useful in explaining consumers’ attitudes and intentions toward a given behaviour. TPB is a general theory of human behaviour while TAM is specific to information systems.

Studies on acceptance of new technology indicate that traditional adoption models need to be extended and modified to better explain the adoption of the innovations. This study modifies TAM by proposing additional acceptance factors associated with mobile services, Rogers (1995) and Figge (2004) having especially influence on the creation of construct items stated in Figure 1. Attitudinal construct was adopted from TPB (e.g. Karjaluoto and Alatalo, 2007). Mobile service adoption model was decided to be based on utilitarian-oriented aspects. Thus, the pure entertainment factors, such as perceived enjoyment and playfulness were not included as construct items. Figure 1 shows the created conceptual model for case mobile service with causal links between the construct items (Ervasti, 2007).

![Conceptual model for mobile services](image)

Perceived usefulness (PU), perceived ease of use (PEOU) and barriers to use (BU) are posited as antecedents of attitude (ATT) toward mobile services. Barriers to use are also conceptualised with context (CON) as an antecedent of intention (INT) to use mobile services. Perceived behavioural control (PBC), reference group influence (REF) and attitude are antecedents of intention, which, in turn, affects actual use (USE). (Fishbein and Ajzen, 1975; Davis, 1989; Ajzen, 1991; Moore and Benbasat, 1991; Thompson et al. 1991; Davis et al., 1992; Taylor and Todd, 1995; Venkatesh and Davis, 2000).

Based on the conceptual model in Figure 1, the following hypotheses for mobile services were created:

- H1: Perceived usefulness of mobile services positively affects attitude toward mobile services.
- H2: Perceived ease of use of mobile services positively affects attitude.
- H3a: Perceived barriers to use mobile services negatively affect attitude.
- H3b: Perceived barriers to use mobile services negatively affect intention to use mobile services.
- H4: Utilisation of context-specific information on mobile services positively affects intention.
- H5: Perceived behavioural control of mobile services positively affects intention.
- H6: Reference group influence positively affects intention.
- H7: Positive attitude toward mobile services positively affects intention to use mobile services.
- H8: Intention to use mobile services is positively associated with actual use of mobile services.

In order to further evaluate and develop the conceptual model with related hypotheses, the use and adoption were tested in a real-life case study of mobile service.

3 THE EVALUATION OF CASE MOBILE SERVICE

The practical evaluation of the mobile service adoption was made in the case study, named Mora. The case mobile service focused on additional value of mobility, which means the possibility to access, produce and send information, products and services at anytime and from anywhere, regardless of the user’s location or the time of day. The major aim of the mobile service was to provide essential information for the end-users as the administration of the campus area defined. Mora service was
evaluated in a delimited user group of students and personnel in Finnish campus area. Mora was launched at the campus, focused on technology and economics education, in Finland in September 2007, in order to research the adoption of mobile services among campus’ students and personnel.

The technical implementation of the case mobile service is based on a client-server approach to providing a mobile access to intranet of campus area. The contents of Mora were internal news, events, announcements and schedules of the study community, as well as a mobile flea market. Mobile flea market enabled users to sell and buy things such as studying material easily and quickly. In later development the mobile service could also provide pleasure, fun, and enjoyment. The mobile client Mora works as an interaction channel between customer and service provider by providing personalised information to the user (Ervasti, 2007). The server side application carries most of the functionality. The client side of Mora is Java mobile application (J2ME) that works in most common mobile phones.

The main technical building blocks of a personalised mobile service concept are the service provider’s existing information systems, mobile service, customer database and mobile application. The main building blocks (adapted from Alahuhta et al., 2005) of the Mora service concept are introduced in Figure 2:

![Figure 2: Overview of the Mora mobile service.](image)

The major source of information for most of the data and communication processed in the Mora service is the existing information system, the campus’ web intranet. This required adaptation of the content from web-based content to mobile content. The user database is utilised for providing personalised services for customers. The existing database is used as the basic information source for user-related information (name, email address, study group), besides the basic personal data the database could be contain more detailed information on customer’s special preferences. The mobile end of the concept utilised mobile Java technologies. The concept approach of using mobile application has several benefits both for the end user and the service provider. Java application provides versatile content delivery and interaction channel, a service provider can create a fully branded mobile service concept and the application is unobtrusive and convenient to use, avoiding the spam effect. When compared to browsing, the application-based approach offers better possibilities for improving usability in mobile services. In addition, the optimisation of data traffic between the terminal and servers is easier to implement in mobile applications. The small size of the displays also favors application-type services.

The case mobile service started with an information and marketing campaign advertising Mora web pages and giving presentations about the field trial to various student groups and personnel in the campus area. The purpose was to make the test-users feel they were privileged to have the mobile service like Mora exclusively on their campus, and have the opportunity to participate in the service testing and development. These initial Mora users were encouraged to compete for the best development ideas. In the adoption of application-based mobile services, the key barrier to overcome is to get the users to do the first download (Nokia, 2003). Therefore, Mora instructions presented simple steps for starting the use of Mora as follows: 1) Registration via web page; 2) Downloading the Mora mobile application via SMS or web page; 3) Setting up the required connection settings; 4) Setting up the mobile client preferences.

The test-users were able to post questions for us during the whole field trial and separate events for getting help in downloading and installing the application were arranged to further ease the adoption process. Users had also an opportunity to test the Mora service with separate test phones in case a user’s own mobile phone didn’t support the application. The nature of the marketing campaign was selected due to the time and costs limitations. In future research the diffusion of mobile services by viral marketing should also be considered and researched.
4 FEEDBACK QUESTIONNAIRE AND ANALYSIS OF THE MORA MOBILE SERVICE

The feedback from the Mora mobile service was collected after two months’ field trial through online and paper questionnaires. The feedback questionnaire was created based on the hypotheses in conceptual model (Figure 1). All of the questionnaire’s construct items consist of multiple variables, and a five-point Likert scale ranging from 1 (strongly disagree) to 5 (strongly agree), with 3 (cannot determine) as a neutral anchor point was used to measure the variables. In Table 1 is presented the feedback questionnaire for the case study (adapted from Venkatesh et al., 2003).

During the field trial, the Mora mobile service gathered 67 registered users and the survey resulted in 52 completed feedback forms. The results were treated and analysed anonymously (Ervasti, 2007). The analysis of the research data was done by using SPSS 14.0 software. A small number of questionnaire forms were incompletely filled in and the answers were ambiguous in some places. The ambiguity was resolved by interpreting the answer in the least favourable way for the question presented.

The rather small amount of sample data restricted the number of feasible analysis methods. The analysis was conducted using the basic frequency and descriptive calculations. The regression analysis was utilised by first examining the effect of the independent variables on the dependent variables according to the set hypotheses, and then testing the combined effect of the explanatory variables as stated in the conceptual model.

Of the respondents, 11.5% were female and 88.5% male and the majority were less than 25 years old. By nationality, 69.2% of the sample were Finnish and 30.8% foreign; the foreign respondents’ nationalities mostly representing Nigeria, Russia and Hungary. Most of the survey respondents had a monthly income of less than 1,000 euros and a relatively new mobile phone in use, the handset’s age being under two years for the majority. The sample’s monthly mobile phone expenditure was mainly under 30 euros, and most used their phone more for private than business purposes.

The survey respondents’ attitudinal valuations of the Mora mobile service are presented in Table 1 with mean values and standard deviations. In perceived usefulness, the respondents mostly agreed with the statement “Using Mora gives me topical information”. In perceived ease of use, most agreed with the statement “I quickly learned to use Mora”, and they were also aware of the kind of mobile services they can use with their mobile phones. Most users agreed with the statement that Mora downloading and installation was easy. Users disagreed with the statements concerning perceived barriers to use. Fear of technical features, required effort in downloading and installing the application, or even the data transmission billing weren’t seen as obstacles in adopting mobile services. The utilisation of time and date-related contextual information in Mora services was seen as more useful than the utilisation of location-specific information. The respondents weren’t especially eager to spend time providing their personal details to make the content in Mora more relevant to their needs.

When examining the perceived behavioural control, the respondents found it important that they could control the use of Mora by deciding when and where to use the service. The reference group influence was perceived as rather irrelevant in this survey. The attitudinal construct showed that the respondents had a more preferable attitude toward downloadable applications than toward SMS-based mobile services. Most agreed with the statement “Using the mobile phone as a channel to access the content of school intra is a good idea”. The respondents reported that they were willing to use

<table>
<thead>
<tr>
<th>Construct item</th>
<th>Variable</th>
<th>Mean</th>
<th>s.d.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Perceived usefulness</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PU1</td>
<td>Using Mora gives me topical information.</td>
<td>4.29</td>
<td>0.610</td>
</tr>
<tr>
<td>PU2</td>
<td>Using Mora saves me time / helps me accomplish things more quickly.</td>
<td>3.77</td>
<td>1.022</td>
</tr>
<tr>
<td>PU3</td>
<td>Using Mora gives me personal information.</td>
<td>3.39</td>
<td>1.150</td>
</tr>
<tr>
<td>PU4</td>
<td>Using Mora gives me information that is of interest to me.</td>
<td>3.81</td>
<td>0.817</td>
</tr>
<tr>
<td>PU5</td>
<td>I find Mora useful in my daily life</td>
<td>3.90</td>
<td>1.053</td>
</tr>
<tr>
<td>Perceived ease of use</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PEOU1</td>
<td>I know what kind of mobile services I can use with my phone.</td>
<td>4.14</td>
<td>1.167</td>
</tr>
</tbody>
</table>
Table 1: Feedback questionnaire items and attitudinal dimensions of test-user responses to Mora mobile service (cont.).

<table>
<thead>
<tr>
<th>Item</th>
<th>Description</th>
<th>Mean</th>
<th>Std. Dev.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>PEOU2</strong></td>
<td>It was easy for me to download and install Mora in my mobile phone.</td>
<td>3.88</td>
<td>1.211</td>
</tr>
<tr>
<td><strong>PEOU3</strong></td>
<td>I quickly learned to use Mora.</td>
<td>4.62</td>
<td>0.718</td>
</tr>
<tr>
<td><strong>PEOU4</strong></td>
<td>I find Mora easy to use.</td>
<td>4.33</td>
<td>0.964</td>
</tr>
<tr>
<td><strong>PEOU5</strong></td>
<td>My interaction with Mora is clear and understandable.</td>
<td>4.00</td>
<td>1.085</td>
</tr>
<tr>
<td><strong>PEOU6</strong></td>
<td>I find it easy to get Mora to do what I want to do.</td>
<td>3.92</td>
<td>1.100</td>
</tr>
<tr>
<td><strong>Barriers to use</strong></td>
<td>Fear of technical features is an obstacle to me adopting mobile services.</td>
<td>1.63</td>
<td>1.085</td>
</tr>
<tr>
<td><strong>BU1</strong></td>
<td>Difficulty and effort caused by the downloading and installation of the application in a mobile phone is an obstacle to me adopting mobile services.</td>
<td>1.76</td>
<td>1.205</td>
</tr>
<tr>
<td><strong>BU3</strong></td>
<td>Knowledge of data transmission billing is an obstacle to me adopting mobile services.</td>
<td>2.51</td>
<td>1.528</td>
</tr>
<tr>
<td><strong>Context</strong></td>
<td>I would view the content of Mora related to a specific time or date (e.g. exam day, weekend) as useful.</td>
<td>4.39</td>
<td>0.750</td>
</tr>
<tr>
<td><strong>CON1</strong></td>
<td>I would view the content of Mora related to me being in a specific location (e.g. campus restaurant, library) as useful.</td>
<td>4.10</td>
<td>0.953</td>
</tr>
<tr>
<td><strong>CON3</strong></td>
<td>I would be prepared to spend time providing my personal details (a user profile) to make the content of Mora more relevant to my needs.</td>
<td>3.78</td>
<td>1.112</td>
</tr>
<tr>
<td><strong>Perceived behavioural control</strong></td>
<td>I find it important that I can choose between different styles in Mora.</td>
<td>3.17</td>
<td>1.354</td>
</tr>
<tr>
<td><strong>PBC1</strong></td>
<td>I find it important that I can easily control the use of Mora (that I can decide when and where I use Mora).</td>
<td>4.50</td>
<td>0.828</td>
</tr>
<tr>
<td><strong>PBC3</strong></td>
<td>I find it important that I can easily stop using Mora.</td>
<td>4.44</td>
<td>0.978</td>
</tr>
<tr>
<td><strong>PBC4</strong></td>
<td>It would be important for me to be in control in terms of the ability to filter the content of Mora related to my interests/needs.</td>
<td>4.19</td>
<td>0.793</td>
</tr>
<tr>
<td><strong>Reference group influence</strong></td>
<td>I started to use Mora because it was recommended to me by someone I know.</td>
<td>2.71</td>
<td>1.576</td>
</tr>
<tr>
<td><strong>REF1</strong></td>
<td>I started to use Mora because my friends are using it.</td>
<td>2.12</td>
<td>1.199</td>
</tr>
<tr>
<td><strong>REF3</strong></td>
<td>I would have been more likely to start to use Mora if the message containing a link for downloading the application was forwarded to me by someone I know.</td>
<td>2.83</td>
<td>1.424</td>
</tr>
<tr>
<td><strong>REF4</strong></td>
<td>I recommended Mora to people I know.</td>
<td>3.49</td>
<td>1.271</td>
</tr>
<tr>
<td><strong>Attitude</strong></td>
<td>Using mobile services (downloadable applications) is pleasant.</td>
<td>3.90</td>
<td>1.159</td>
</tr>
<tr>
<td><strong>ATT1</strong></td>
<td>Using Mora is pleasant.</td>
<td>4.08</td>
<td>0.710</td>
</tr>
<tr>
<td><strong>ATT3</strong></td>
<td>Using mobile services (SMS-based) is pleasant.</td>
<td>3.38</td>
<td>1.223</td>
</tr>
<tr>
<td><strong>ATT4</strong></td>
<td>Using mobile technology to access a variety of services is interesting.</td>
<td>4.31</td>
<td>0.707</td>
</tr>
<tr>
<td><strong>ATT5</strong></td>
<td>Using the mobile phone as a channel to access the content of the university intranet is a good idea.</td>
<td>4.73</td>
<td>0.528</td>
</tr>
<tr>
<td><strong>Intention</strong></td>
<td>I feel positively about Mora.</td>
<td>4.54</td>
<td>0.503</td>
</tr>
<tr>
<td><strong>INT1</strong></td>
<td>I am willing to use Mora in the future, if possible.</td>
<td>4.48</td>
<td>0.641</td>
</tr>
<tr>
<td><strong>INT3</strong></td>
<td>I would use Mora regularly in the future.</td>
<td>4.02</td>
<td>1.038</td>
</tr>
<tr>
<td><strong>Actual use</strong></td>
<td>I am using Mora every day.</td>
<td>2.58</td>
<td>1.226</td>
</tr>
<tr>
<td><strong>USE1</strong></td>
<td>I am using Mora every week.</td>
<td>3.71</td>
<td>1.226</td>
</tr>
<tr>
<td><strong>USE3</strong></td>
<td>I regularly check what’s new with Mora.</td>
<td>3.40</td>
<td>1.302</td>
</tr>
</tbody>
</table>
Mora in the future. However, the actual use of Mora wasn’t that regular since it happened more on a weekly than daily basis.

71.2% of respondents reported having used other mobile services in addition to Mora and 67.3% of them had used downloadable applications, whereas only 38.5% SMS-based services. 57.7% of respondents reported having used entertainment services, and 55.8% had used both the information and news services and email; 38.5% had used instant messaging, and 23% banking and financial services. 17.3% of respondents had experience of VoIP, 11.5% of ticket reservation, 7.7% of shopping, and 5.8% of travel booking.

Table 2 shows the standardised coefficients and t-values of the tested hypotheses regarding mobile services (t-statistics indicating the significance of individual variables when the t-value is > 2 (Schwager, 1995).

<table>
<thead>
<tr>
<th>Hypothesis</th>
<th>Standardized Coefficient</th>
<th>t-value</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>H1: PU to ATT</td>
<td>.419</td>
<td>3.164</td>
<td>.003</td>
</tr>
<tr>
<td>H2: PEOU to ATT</td>
<td>.303</td>
<td>2.182</td>
<td>.034</td>
</tr>
<tr>
<td>H3a: BU to ATT</td>
<td>-.173</td>
<td>-1.205</td>
<td>.234</td>
</tr>
<tr>
<td>H3b: BU to INT</td>
<td>-.050</td>
<td>-0.345</td>
<td>.731</td>
</tr>
<tr>
<td>H4: CON to INT</td>
<td>.539</td>
<td>4.430</td>
<td>.000</td>
</tr>
<tr>
<td>H5: PBC to INT</td>
<td>.250</td>
<td>1.828</td>
<td>.073</td>
</tr>
<tr>
<td>H6: REF to INT</td>
<td>.237</td>
<td>1.707</td>
<td>.094</td>
</tr>
<tr>
<td>H7: ATT to INT</td>
<td>.572</td>
<td>4.882</td>
<td>.000</td>
</tr>
<tr>
<td>H8: INT to USE</td>
<td>.381</td>
<td>2.913</td>
<td>.005</td>
</tr>
</tbody>
</table>

The strong coefficient of H1 shows that there is a positive and direct relationship between PU and ATT. Thus, the perceived usefulness of mobile services is a strong predictor of attitude and H1 was supported. It can also be seen that the PEOU of mobile services was directly and positively associated with ATT, providing support for H2. H3a and H3b address the role of BU, the coefficients were negative, but very low, which indicates that barriers to use have no significant influence on either attitude or intention, thus H3a and H3b were not supported. Next, the strong coefficient of H4 indicates that context-related information acts as a strong and positive predictor of intention to use mobile services, so H4 was supported. H5 and H6 hypothesise that PBC and REF have a positive and direct relationship with INT. Both coefficients are similar with low magnitudes, therefore neither H5 nor H6 were supported by the data. H7 states that the more favourable attitude a person forms toward mobile services, the higher his or her intention to use those services. The relationship was found to be positive and statistically significant. The final hypothesis, H8, argues that INT is directly associated with USE, and the coefficient is strong and positive. Thus both H7 and H8 were supported.

Figure 3 illustrates the standardised coefficients for the conceptual model.

During the field trial it was noticed that mobility does not provide enough added value for the users when the Mora service is in its current form. Consumers were not attracted to the service because they already had an easy and familiar way to access almost the same content via the web, so they felt that the Mora is just a mobile substitute for the campus’ web intranet. Even though the challenges with marketing the Mora service and establishing test-users’ interest were recognised and actions were taken to overcome them, one reason for the somewhat modest use of the Mora service was that the benefits of the service weren’t enough visible to potential adopters. For Mora service users, the easiness and flexibility of the service should have been stressed even more, i.e. that service can be used at any time and in any place, freeing the users from the traditional time and place constraints even more efficiently than Internet services. Another significant obstacle was the effort required in downloading and installing the application. To diminish this problem several separate events were arranged where guidance and support in service adoption was offered to users.

5 CONCLUSIONS

This paper introduces a conceptual model for mobile service adoption that is a modified version of established adoption models. The developed conceptual model was evaluated and validated in the
real-life case study of mobile service, through which we have gained practical insights into application-based mobile service development and promotion, as well as having identified the factors affecting the adoption of mobile services in Finland.

The findings from the survey suggest that perceived usefulness of mobile services is the strongest predictor of attitude toward mobile services. That is also supported by previous studies (e.g., Rogers, 1995) on consumer acceptance of new technologies. People will adopt a technology when the added value offered by technology fulfills their needs. Hence, compatibility with users’ needs (ibid) is a critical variable for predicting technology adoption. Ease of use wasn’t perceived as having a strong relationship with acceptance of mobile services, contrary to (ibid), so it can be reasoned that usefulness is a stronger determinant of attitude toward mobile services than perceived ease of use. Utilisation of contextual information was found to have a strong impact on intention to use mobile services and proves that consumers desire relevant and personalised services.

The respondents did not perceive the technical features of mobile services, the downloading and installation process or data transmission billing as significant barriers to adopting mobile services as prior researches suggest (Aarnio et al., 2002; Nokia, 2003). When examining the impact of control and social norms, the results of the survey were not congruent with the theoretical background (Ajzen, 1991; Rogers, 1995) either, since perceived behavioural control and reference group influence weren’t stated as important determinants of service adoption. A preferable attitude toward mobile services was however proved to have a strong impact on intention to adopt services, whereas the intention in the end didn’t act as a good predictor of actual use.

Based on the case study findings, it is suggested that the extended and modified conceptual model can be used to guide industry players’ evaluation of the adoption potential of new mobile services. Case study results can be used to help mobile service providers choose right development and marketing strategies for mobile services and accelerate the diffusion of services. However, further and more extensive studies should be conducted in order to attain a more profound understanding of the issues concerning consumers’ drivers for adopting (or rejecting) mobile services. Similar research need to be done in other countries as well in order to obtain benchmarks for comparing the differences in adoption characteristics in different countries, and to further validate the conceptual model. In an international comparison, Finland is an advanced information society especially when it comes to mobile communication. Thus, the study findings might vary in other countries with far lower mobile phone penetration rates and different mobile communication culture. However, when compared to Japan, Finland is far behind in the versatility of mobile communication: in Japan, a nation of 127 million the number of mobile internet subscribers recently passed 100 million (The Guardian, 2007). The conceptual model should also be extended further to include more factors that reflect the unique characteristics of the usage intentions of mobile services. Further attention should be devoted to the comparison of rival models in explaining consumer attitudes and intentions.

REFERENCES


SHORT PAPERS
THE ROLE OF AUDIO-VISUAL METAPHORS IN AIDING THE COMMUNICATION OF CUSTOMER KNOWLEDGE

User Satisfaction Perspective

Dimitrios I. Rigas and Mutlaq B. Alotaibi
Department of computing, University of Bradford, Bradford, West Yorkshire, BD7 1DP, U.K.
d.rigas@bradford.ac.uk, m.b.g.alotaibi@bradford.ac.uk

Keywords: Customer, Knowledge, Multimodal, Interaction, Satisfaction, Sound, Speech, Earcons, Auditory Icons, Audio-visual.

Abstract: The purpose of this paper was to examine the implications of employing multimodal user interaction in Electronic Customer Knowledge Management Systems (E-CKMS). There are various challenges to E-CKMS mentioned in current literature and reported (trust and knowledge hoarding, structured transfer of knowledge and content and relevancy issues). As empirical CKM studies that examine the role of multimodal interaction in this field are generally lacking, this paper reports research conducted to evaluate this role and measure user satisfaction. This evaluation was undertaken empirically by developing two E-CKMS experimental platforms (text with graphics only and multimodal). The major findings indicated that user satisfaction was significantly improved by using multimodal metaphors. Further investigation is needed to determine the usefulness of two multimodal versions on E-CKMS instead of comparing text with graphics only with multimodal.

1 INTRODUCTION

The new means of IT, such as internet, extranets and intranets, leads to dramatic shift from information to knowledge societies (Thierauf, 1999, Goh, 2005). Harnessing invisible assets is becoming one of the primary sources of creating value and competitive advantage in the current age of knowledge (Goh, 2005). Knowledge as a concept covers vast area of various views (Becerra-Fernandez et al., 2004), levels (Davenport and Prusak, 1998, Rowley, 2002), principles (Such et al., 2001), taxonomies (Davenport and Prusak, 1998, Thierauf, 1999, Hahn and Subramani, 2000), strategies and trends. One classification of knowledge is based on the source from which knowledge has been elicited (external and internal) (Davenport and Prusak, 1998). External knowledge (CK) can be regarded as one of the most organisations valuable types of knowledge (Osterle, 1995), which is gathered at the customer point of contact under a great deal of time pressure (Lesser et al., 2000). The Utilisation of CK helps organisations to grow, innovate and compete against competitors (Gebert et al., 2002a, Gebert et al., 2002b), but it is not easy to gather, identify, interpret and integrate, because it flows to the organisation form multiple communication channels (Bueren, 2005). This led to integrating both Knowledge Management (KM) and Customer Relationship Management (CRM) in E-Business contexts. More details on the determination and characteristics of KM, CRM and E-Business can be found in (Becerra-Fernandez et al., 2004, Tiwana, 2001, Skyrme, 2001, Alotaibi and Alzahrani, 2004).

E-CKMS is derived from the integration between KM and CRM in E-Business (Tiwana, 2001), and its typical example is represented in Amazon case study (Gebert et al., 2002b, Gurgul et al., 2002, Rollins and Halinen, 2005), which illustrates how Communities of Customers (CoC) works. There are several similarities between CoC and Communities of Practice (CoP), which mean that CoC is rooted in the traditional KM (Lesser et al., 2000, Gurgul et al., 2002). In CoC, peer customers share opinions and insights about products and services, which can lead to more effective and guided decisions made by any member of this community (Gibbert et al., 2002). From CRM point of view, storing and analysing customer historical transactions leads to understanding customer buying patterns, and hence leveraging up-selling and cross-selling opportunities (Tiwana, 2001, Pan and Lee, 2003). Amazon case study involves KM and CRM aspects that can be
benchmarking by efforts devoted to evaluating E-

CRMS. Interactive systems can be seen as one of the

CRM components that enable E-CRM by

establishing learning relationship (learn while

interact) (Massey et al., 2001). Interacting with

customers in real-time and adopting CoC can be

counted among several proposed approaches that

improve the elicitation (Gibbert et al., 2002), expand

the exploitation (Lesser et al., 2000) of CK, reduce

cost, and hence replace the traditional approach that

rely on understanding of sales representatives or

results market research.

There were several challenges to E-CKMS that

can be tackled using the interactive multimodal

metaphors. Knowledge hoarding (Davenport and

Prusak, 1998, Gibbert et al., 2002), for example, can

be seen as the lack of customer willingness to share

knowledge, and solved by several approaches that

include optimising customer-company dialogue

(Interaction) (Massey et al., 2001, Gurgul et al.,


stated that trust and knowledge hoarding could be

tackled by not only establishing continuous two-way

dialog with customers, but also employing

interactive multimedia systems. Another challenge is

that customers when interacting with E-CKMS

needs to have their ideas well-structured and

organised, which can be aided by what so-called

Customer Innovation Toolkits (CIT) (von Hippel,

2001b, Von Hippel, 2001a). CIT can be incorporated

into E-CKMS in order to facilitate an optimal

transformation of customer expertise and

expectations (knowledge sharing) into valuable

suggestions, which can be afterward used to offer

customised and personalised products and services.

This context involves knowledge sharing, and hence

encouraging customer to share knowledge is needed,

which leads to the assumption that there is a

potential role for multimodal interaction metaphors
to play. Another challenge is information overload

and relevancy, which is, actually derived the web-

based environment. Authors (Davenport and

Prusak, 1998, Bueren, 2005) in E-CKMS field raises

information overload and relevancy as a concern,

and include it with content issues, but some mention

identification of relevant knowledge (Urban and von

Hippel, 1988). Brewster (1997) argued that this
could be addressed by enhancing the text with

graphics only manner of information display with

means of auditory metaphors (Brewster, 1997). To

sum up, multimodal interaction is anticipated to

dress trust, structure and content questions in E-

CKMS.

The remainder of the paper is organised in seven

sections. In Section 2, we introduced relevant work.

Section 3 described the experimental platform.

Design of the empirical study is shown in Section 4.

In Section 5, we presented results and discussion.

Conclusion is provided in Section 6. Finally, we
described future work in Section 7.

2 RELEVANT WORK

Researchers in the CKM field tend to analyse CKM

case studies, identify cultural, structural and

managerial mechanisms that facilitate CKM success

and propose models and frameworks for theory and


proposed a model of social relationships, and put

emphasis on interpreting knowledge directly by

human (García-Murillo and Annabi, 2002). Some

other authors consider structural aspects, such as

organisational boundaries, culture, structure and

climate (Bose and Sugumaran, 2003, Dous et al.,

2005, Bueren, 2005). Others look at rewards

systems, incentives, management support and

perception (Gibbert et al., 2002, Gurgul et al., 2002).

In addition, there were several authors who
discussed CK characteristics, approaches,

applications and relations (Feng and Tian, 2005,


Business Engineering (BE) (Osterle, 1995)
perspective suggests separating strategy, process,
system and change levels (Bose and Sugumaran,

2003, Dous et al., 2005, Bueren, 2005). Several

studies presented styles of CKM and types of CK

(Gibbert et al., 2002, Gurgul et al., 2002, Feng and

Tian, 2005, Rowley, 2002). Although, these studies

commonly studied CK based on several perspectives

and points of view, other than empirically examining

the role of multimodal interaction, it provided

insights into the underlying principles and

theoretical foundations of E-CKMS.

There is little known about efforts in CKM field

has been devoted to evaluate the potential role that

multimodal metaphors can play in E-CKMS.

Nevertheless, a great deal of studies in several fields

of study have been conducted to evaluate such role,

and found that user interface can be improved by the

augment of speech (Kehoe and Pitt, 2006)

(synthesised and recorded speech) and non-speech

sounds (earcons (Rigas et al., 2000, Rigas and Alty,

2005) and auditory icons (Gaver, 1997, Cohen,

1993). In software engineering, There were several

studies that support this view, such as (Sonnenwald

et al., 1990, Cohen and Ludwig, 1991, DiGiano et


Overall, user satisfaction and other usability
attributes can be improved in general Information Systems (IS) by employing multimodal interaction.

Table 1: E-CKMS visual and auditory metaphors.

<table>
<thead>
<tr>
<th>CK category</th>
<th>VCKMS</th>
<th>MCKMS</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trends (top 10)</td>
<td>T</td>
<td>G</td>
</tr>
<tr>
<td>Customer review</td>
<td>√</td>
<td></td>
</tr>
<tr>
<td>Customer rating</td>
<td>√</td>
<td></td>
</tr>
<tr>
<td>Website advice</td>
<td>√</td>
<td></td>
</tr>
<tr>
<td>Co-production</td>
<td>√</td>
<td></td>
</tr>
<tr>
<td>Product features</td>
<td>√</td>
<td></td>
</tr>
</tbody>
</table>

Burke et al. (2006) carried out meta-analysis investigation into the effectiveness of multimodal interaction, in forty three studies, and found that audio-visual metaphors have a significant role to play in improving user performance in IS compared to the visual-only display. In addition, this finding was supported by the results obtained from two experiments carried out by (Rigas and Memery, 2002), who investigated the use of auditory stimuli (speech, earcons and auditory icons) to communicate information to users in both email and stock control applications. Additionally, two studies (Rigas and Alty, 2005, Alty and Rigas, 2005) investigated utilising rising pitch metaphors in the communication of graphical information, and found that it was possible for visually-impaired users to interpret graphical information with the aid of rising pitch metaphors, even in the absence of a visual display.

Since E-CKMS is a web-based environment, it can be linked to similar fields of study, such as web-based browsing and email applications (Rigas and Memery, 2002, Rigas and Memery, 2003, Rigas, 2003). In web-based browsing systems, a prototype has been develop as an online help system with sound support (Kehoe and Pitt, 2006), and extended later by incorporating no-speech sound and other auditory metaphors (Kehoe et al., 2007). In addition, another web-based browsing prototype was built to browse musical notes with the help of sound, and proven successful performance improvement (Fernström and McNamara, 2005). In email applications, several experiments (Rigas and Memery, 2002, Rigas and Memery, 2003, Rigas, 2003) were conducted to evaluate the potential of audio-visual metaphors in reducing visual complexity and tackling information hiding, and concluded that this hypothesis was true, besides that visual display have to be synchronised with means of auditory stimuli.

3 EXPERIMENTAL PLATFORM

The experimental platform developed for this research provided typical functions of web-based mobile phones retailing systems, and included an additional function labelled as co-production, which defined by (Gibbert et al., 2002) as the manner in which customers practice New Product Development (NPD). Typically, E-CKMS consists of three main components: CKM, infrastructural and user interface components, and can incorporate any additional function, such as the one included in this study (co-production). This study assumed that CKM and infrastructural components were previously implemented.

Co-production function facilitates the dual role of customers (producer and consumer) by offering CIT that aid customers to manipulate elements in solution space to test new products (billing scheme). In fact, solution space included several elements, but the scope of this study limited these elements to loyalty, billing and taffies schemes. Furthermore, co-production function offered a trail-and-error engine that enabled experimental NPD, allowed its repetition until final product design was reached. This engine received customised schemes from CIT, sent it to billing engine, received customised bill, stored it in trails comparison array and provided comparison of results obtained from other trails in order to support customer decision making.

This platform was implemented with two interfaces. These two versions were text with graphics only E-CKMS (VCKMS) and Multimodal E-CKMS (MCKMS). In VCKMS, CK was communicated to E-CKMS users using text with graphics only, while the communication method was audio-visual in MCKMS. The communication of CK required classification of CK and auditory and visual metaphors, and utilisation of a wide range of technologies. First, Types of CK were organised into six categories (trends, customer reviews, customer ratings, website advices, co-production CK and product features). Second, the visual metaphors employed were text (T) and Graphic (G), whilst the auditory metaphors were synthesised speech (S), earcons (E), auditory icons (A) and recoded speech (R). Table 1 shows each CK category and the way by which it was communicated.
4 DESIGN OF THE EMPIRICAL STUDY

The research undertook explored customer satisfaction aspects related to the use of audio-visual metaphors in three levels of task complexity.

Table 2: Summary of task levels influential factors.

<table>
<thead>
<tr>
<th>Task</th>
<th>CKMA Complexity factors</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>PSA TSA PRA CI CI NR NS</td>
</tr>
<tr>
<td>E</td>
<td>A1 A2 L L 10 40</td>
</tr>
<tr>
<td>M</td>
<td>A3 A4 A5 M M 15 17</td>
</tr>
<tr>
<td>D</td>
<td>A6 A7 A8 H H 17 4</td>
</tr>
</tbody>
</table>

The three levels are task easy (Task E), task moderate (Task M) and task difficult (Task D). In this study, task levels were created based on six influential factors: number of task requirements (NOTR), number of available selections (NOAS), number of CKM activities (CKMA), customer interaction (CI) and CK intensity (CKI). If the task was to be designed as difficult, NOTR, CKMA, CKI and CI needed to be increased, while NOAS was to be decreased. In CKMA factor, there have been a various types of CKMA: phone selection activity (PSA), tariff selection activity (TSA) and Co-production activity (CPR). In CI and CKI factors, there were three important levels: low (L), moderate (M) and high (H). Table 2 reviews the task complexity influential factors, and illustrates the association between tasks and CKMA. More information on task levels, types and workload is provided on (Burke et al., 2006).

Forty subjects (all were students at University of Bradford, and regular internet users) were selected randomly, based on the non-probability sampling strategy (convenience-sampling method) (Salkind, 2006b). Subjects were divided into two groups (20 each): control and experimental, and then offered a short training session on the corresponding version of E-CKMS. Subsequently, subjects were asked to perform the three tasks and then fill a questionnaire devised for this study. The order of tasks was counterbalanced between participants in order to neutralize possible task learning effect.

5 DISCUSSION OF RESULTS

Satisfaction was measured by a set of user provided answers to questionnaire questions, which include ease of the system (EOS), extent of user confusion (EOC), extent of user frustration (EOF), ease of navigation (EON) and overall comfort (COM). Measuring user attitude towards the system appears to be difficult. However, asking to specify the extent, to which the user agree or disagree with a set of statements, tend to support the pursuit of this measurement (Jordan, 1998).

Table 3: The mode and frequency values of the five aspects of customer satisfaction.

<table>
<thead>
<tr>
<th>Group</th>
<th>Value</th>
<th>Aspects of customer satisfaction</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Mode</td>
<td>EOS  EOC  EOF  EON  COM</td>
</tr>
<tr>
<td>VCKMS</td>
<td>Mode</td>
<td>4     3     3     5     5</td>
</tr>
<tr>
<td></td>
<td>Freq.</td>
<td>50%   40%  45%  85%  55%</td>
</tr>
<tr>
<td>MCKMS</td>
<td>Mode</td>
<td>5     2     2     5     5</td>
</tr>
<tr>
<td></td>
<td>Freq.</td>
<td>70%   55%  50%  65%  65%</td>
</tr>
</tbody>
</table>

User agreement and disagreement utilised a six-point scale ranging from agree strongly to disagree strongly (Salkind, 2006a). The values of the scale were six for strong agreement, five for moderate agreement, four for slight agreement, three for slight disagreement, two for moderate disagreement and one for strong disagreement. After completing all user satisfaction questions, responses were summed up to generate an overall score for user satisfaction, based on the system usability scale (SUS) (Brooke, 1996).

It was noteworthy that multimodal interaction reduced customer response time, which led to greater customer satisfaction and eventually loyalty. Participants in the experimental group expressed interest in CK communicated aurally more than those in the control one. Our experience with this platform suggested that users tend to be more comfortable with aural communication when sounds conveyed more rapidly than for the first time. At first glance, the mean value of customer satisfaction for using MCKMS (77%) was higher than that for VCKMS (63%). Significance of the difference between the two conditions was tested (at 0.05 significance level) using the t-test. The difference was found significant (t38 = 4, CV= 2.03 P < 0.05).

Table 3 shows the mode values for the aspects of user satisfaction with the values for using VCKMS and MCKMS, in addition to the frequency of the mode. User responses suggested that the multimodal E-CKMS is easier to use, less confusing and less frustrating. In fact, 70% of users agreed moderately that the multimodal system was easy to use compared to 50% agreed slightly that the text with graphics only version was easy to use. In user confusion and frustration, half of the sample agreed
moderately that the multimodal E-CKMS is neither confusing nor frustrating, in comparison to 40% and 45% of users disagree slightly that the text with graphics only version was confusing and frustrating respectively. Furthermore, Mann-Whitney statistical test was performed, and showed that there were significant differences in all aspect of customer satisfaction expect in the ease of navigation.

6 CONCLUSIONS

The role E-CKM has become increasingly important to public and private organisations due to the cost savings it offers. However, it is considerably complex, vague and challenging discipline due to the many aspects involved. This study shed light into three challenges to E-CKMS (trust and knowledge hoarding, structured knowledge transformation and content and relevancy issues), which can be tackled by utilising multimodal metaphors. This hypothesis was tested empirically by two independent groups in two E-CKMS. Subjects took part in this research were satisfied and expressed interest in the use of audio-visual metaphors. The significance of this approach was found in all aspects of customer satisfaction, except in the ease of navigation. Therefore, it can be concluded that audio-visual metaphors has contributed positively towards the improvement of customer satisfaction. In summary, results showed that the use of multimodal interactive metaphors in E-CKMS is more satisfactory than to text with graphics display.

7 FURTHER WORK

This experiment reported here proved to be successful. However, it was noticed during the previous experiments that users tend not to take into account various types of presented CK, such as customer ratings and reviews. Therefore, further investigation into additional multimodal metaphors is needed to determine how speech with aviator can promotes further utilisation of customer reviews, and this is expected to yield a set of useful guidelines.

REFERENCES


EUROMICRO-97, New Frontiers of Information Technology, 78–84.
A PERVASIVE NUTRITIONAL MONITORING AND ADVISE SYSTEM

NutriMe

Vitor Basto, João Varajão and António Cunha
Department of Engineering, University of trás-os-Montes e Alto Douro, Quinta de Prados, Vila Real, Portugal
vbasto@utad.pt, jvarajao@utad.pt, acunha@utad.pt

Keywords: Nutrition, Pervasive Nutritional Monitoring, Nutritional Advise System.

Abstract: It is well known, widely accepted, scientifically proved and published by major governmental and non-governmental organizations worldwide (e.g. WHO - World Health Organization), that nutritional misbehaviour in so called developed countries, is a major cause of diseases, morbidity and death. The phenomenon is mainly felt in aged populations, but a significant increase has also been detected more recently in young populations. This paper presents a proposal to tackle serious social and behavioural problems related to aging and nutrition. NutriMe is presented as a nutritional monitoring and advising system to help individuals to monitor and correct their behaviours. We also propose NutriMe as the main component for a public national observatory on nutritional profiles for public health analysis purposes.

1 INTRODUCTION

Several economical, social and cultural factors that took place in the last decades in developed countries have strongly influenced human diseases profiles. Among those factors, globalization and urban prevalent and increasing lifestyles are worth to mention. One of the reflexes turning out from those changes is related to nutritional (mis)behaviours (Lopes et al., 2006). Although other reflexes are also subject of research (e.g. smoking, pollution, etc.), we will not take them into account in our study, knowing that nutrition is referred as one of the most important factors.

The outcome of several years of research about nutritional reflexes on health/diseases, lead to a present common sense assumption that proper nutritional monitoring and advise is need, must be continuous, rigorous and customised for each individual, according to biological, medical, and life style parameters (Lopes et al., 2006).

The extent of many harmful reflexes (e.g. morbidity and mortality) caused by incorrect nutritional behaviours on health, have been estimated in several studies. World Health Organization (WHO) reports that 80% of cardiovascular disease cases, 90% of diabetes mellitus type 2 and 33% of all types of cancer could have been prevented by adopting healthier lifestyles, which includes correct nutrition, regular physical activity and non-smoking (WHO, 2006). WHO also states that: “A change in dietary habits, physical activity and tobacco control, have a major impact in reducing the rates of these chronic diseases, often in a relatively short time”.

Nutritional monitoring and advise is therefore important both in an individual point of view for individual behaviours correction, and in a global point of view, essential for global policy definitions and for nutritional education planning.

NutriMe, the system proposed in this paper, is a system designed to tackle the above presented issues. NutriMe can also be integrated with the Smart pantry project - Diet module (Alves et al., 2006), whose features consist of a subset of NutriMe features in the context of a smart house project focused on accessibility and inclusion.

In section 2 a general characterization is made of the Smart pantry project and explained its possible integration with the NutriMe system, followed by section 3 which presents the NutriMe system. Section 4 shows a NutriMe prototype and some issues related to the distributed data model are presented in section 5. Finally, conclusions and future work are referred in section 6.
2 SMART PANTRY PROJECT

Developed countries aged populations is raising several and severe problems related to home daily elder people tasks, such as physical access and handling of home stored goods, nutritional control and health care, etc. The Smart pantry project (Alves et al., 2006) was developed with the intention to fulfil some of those needs. Its main features are related to stock control, triggering advertises to replace depleted/missing products, create automatic customized shop lists, make products physically accessible to people with reduced mobility, suggesting menus tailored to user preferences and medical profiles (diets), all supported by a central database system (Barrias et al., 2008). Figure 1 presents a modular architecture for the smart pantry as proposed in (Alves et al., 2006).

Figure 1: Modular architecture for the smart pantry
Source: (Alves et al., 2006).

In this system, users enter products using the entry module, which automatically identifies the product, presents its features to the user and asks him for confirmation/validation. After validation, the system stores the product according to its storage conditions (temperature and humidity), size, expiry dates, etc.

Based on the storage information in the database, the stock management module manages the existence of stored products (e.g. expiry dates, so that they can be disposed or used), a ‘shopping list’ is also presented based on predicted needs, on-line shopping with supplier’s systems integration (Cardoso et al., 2007), history of menus, etc.

The storage module consists of a robotic storage system, adapted to user needs (accessible) and to available space and conditions. It may be based on rotating shelves, suspended elements, or other systems such those employed in large-scale storage systems.

In order to produce a prototype, two basic solutions are under development: one thought to be part of traditional kitchen furniture (a numeric-control system drawing its inspiration from a bucket chain); another will be a robotic system of the ‘mini-warehouse’ variety, to be installed in its own compartment – the ‘pantry’ (modular systems of greater capacity).

An extremely important element is the HCI (nicely) module (human computer interface module) that will be programmed according to user profiles and specific needs, so that the system can interact with each person according to his personal physical/psychological skills and impairments (image-based, text-based, voice based, simplified vs. full featured versions).

The diet module has a set of menus that allow the counselling of users, accounting for different criteria, such as the number of people (family size), the status of the food stock, or the way for eating adequately.

Smart pantry project diet module features are considered as part of Nutrime project as the home based monitoring and advising features. Nutrime intends to cope with ubiquitous nutritional monitoring also in other contexts like restaurants, canteens, bars and pubs, vending machines (food, beverage, cigars automatic selling machines), etc.

Nutrime extends the concept of nutritional monitoring to ubiquitous individual monitoring and advising system.

3 NUTRIME

NutriMe is a distributed software and distributed data based system that collects nutritional information from different sources. In addition to individual nutritional information gathering, it also links, relates data, ensures consistency and integration of syntactic and semantic data models from different sources (food suppliers – home, restaurants, vending, etc.). NutriMe uses that information not only for individual real-time nutritional monitoring and advising, but also for the purpose of feeding a nutritional national database (nutritional ‘observatory’). The observatory database allows for classification, segmentation and prediction of nutritional profiles. Based on data analysis and knowledge extraction (data mining) from this database, public health policies and strategies can be better supported and deployed. Figure 2 shows the components of NutriMe (conceptual view).
Each component relies on different technologies and targets different purposes and features of a global nutritional monitoring and advising system. *NutriMe* conceptual model, its individual components roles and descriptions, follow:

- **Home desktop component**, supports all the features mentioned in the *Smart pantry project* – *Diet module*. The major features specified are related to user profile characterisation (age, gender, physical activity profile, professional profile, medical profile, etc.), alimentary items characterisation (identification, common description, nutritional data composition, etc.), diets, healthy profiles characterisation, nutritional reports generation and nutritional advise;

- **Mobile device component**, implements a mobile version of the Home desktop component and features for interoperation and integration with all the other system components that provide (or collect) nutritional data for the individual, like the Home desktop, Restaurant, Automatic selling machines and Nutritional observatory components. It is intended to download and run mobile code made available from the other system components (e.g. in restaurants), behave like an extended GUI (graphical user interface) of those systems and interchange (collect/provide) information about individual user consumed items from those system components. The consumed items identified and/or described are to be found in the mobile device database for nutritional monitoring, processing and advise purposes. The mobile component is also able to look through consumed items description provided by the other system components, map them into the nutritional characterisation and run the respective nutritional monitoring, evaluation/classification and advise algorithms;

- **Restaurant component**, our concept of a “smart restaurant” system component includes the following major features: provide mobile applications/code to be downloaded into customer mobile devices allowing for customer multimedia interactive menu selection (e.g. food, drink, desert, etc.), for delivery selection (e.g. tables and location of the customers to be selected based on interactive maps), for customer identification and profile management (e.g. personal data for invoice and receipts), detailed and electronic invoice issuing (detailed info about consumed items) sent to the customer mobile device, electronic payments, etc.;

- **Automatic selling machines (vending) component**, this is a generic component representing any other system component providing and/or collecting nutritional information (e.g. beverage and food automatic selling machines). Each of these components must provide mobile code to be run on user mobile devices or implement a compliant generic communication protocol profile (to be defined) for integration with the mobile device;

- **Nutritional observatory component**, includes features of data collection from individual devices (desktop and mobile devices) and restaurant systems, data synthesis, nutritional profiles analysis based on multiple criteria (e.g. age, location/geography, profession, etc.), allowing for population risk classification and evaluation concerning nutritional behaviours and diseases prevalence analysis.

*NutriMe* intends to promote healthy nutritional behaviours by the means of ubiquitous nutritional monitoring in an individual and population basis (reporting individual and population nutritional warnings). In addition, it is intended to provide detailed individual nutritional advising (suggesting detailed meals according to nutritional principles and user preferences), global behaviour synthesis, risk evaluation and classification using data mining techniques.

Figure 3 shows UML (Unified Modeling Language) Use Cases specification for *NutriMe*.

It shows the system actors, features, their interactions and relationships (Fowler, 2003).
Use Case “Consumer profile setup” allows the user to provide his personal data (e.g. gender, age, height, weight) to the system, defining nutritional profiles, diets and preferences (e.g. vegetarian, athlete, diabetic, etc).

Use Cases “Nutritional misbehaviour alerts” and “Nutritional monitoring reporting and advising” allows for automatic warnings from NutriMe towards the user, allows the user to access statistics about his behaviour and provides nutritional advising. In addition, NutriMe provides extra nutritional expertise and tools for knowledge extraction (clustering, classification, prediction) from monitoring data for nutrition professionals (“Nutritionist”) decision support in “Individual nutritional and health data analysis and data mining” Use Case.

An observatory component of NutriMe is also fed by nutritional monitoring data (Use Case “Nutritional data observatory feed”) for “Group nutritional and health data analysis and data mining” purposes.

“Virtual interfaces setup” Use Case supports user interface, data and functional integration features between NutriMe and “Food supplier systems”. For instance, it is possible for “Food consumers” to issue meal orders from NutriMe that are transmitted electronically and processed by “Restaurant systems”. When “Food consumers” interact with “Food supplier systems” through

NutriMe “Automatic consumption registration” is possible. “Automatic consumption registration” is the preferred way of nutritional data gathering in NutriMe, but “Manual explicit consumption registration” by “Food consumers” are also available in NutriMe.

NutriMe simplified data model is presented in an Entity-Relationship diagram in Figure 4. The conceptual data model will be deployed as a distributed relational database system as detailed in section 5. As a distributed software and distributed data based system that collects nutritional information from different sources and heterogeneous technologies, concerns of consistency, correctness, integration of syntactic and semantic data models are to be considered.

An innovative implementation of NutriMe involving flexible and adaptive information systems integration based on desktop and mobile devices is presented in section 4.

4 PROTOTYPE

Inclusion and accessibility have been major concerns during all development life cycle of NutriMe. Software design and technology selection were essential to construct a solution that promotes device and software mobility, usability, high availability and continuous monitoring.

Mobile technologies such mobile phones/PDAs proved recently to ease cumbersome daily task.
Recent devices reached significant processing storage and communication capabilities, allowing increasing features and new applications. The massive adoption of this kind of devices is considered today one of the greatest success of information and communications technologies market acceptance (Paes, 2006).

The above mentioned technical factors, growing market acceptance, easy of use and ubiquity of this kind of mobile technologies, lead us to select this computing and communication technologies as the basic support for the main NutriMe component.

Figure 5 shows the technological solution proposed for NutriMe system main components, including physical/logical elements, interoperation and interactions.

![Figure 5: Technical scenario for NutriMe mobile component prototype.](image)

NutriMe software (mobile and desktop) components are hosted in PC like and mobile devices (e.g. mobile phones/PDAs).

NutriMe desktop and mobile versions have been developed with Microsoft .NET 2005 Framework and Microsoft .NET Compact Framework respectively. Microsoft SQL Server 2005 and Microsoft SQL Server 2005 Compact were adopted for desktop and mobile support respectively. Data replication, integrity and synchronisation are supported by both mentioned DBMS (master-slave data replication and synchronisation) which is the most common strategy adopted for similar scenarios. This strategy shows optimal trade-offs concerning quality of service (availability, punctuality, etc.) and resources consumption/usage (storage, processing and communication) within strongly connected, weakly connected and connectionless situations. Support of multiple communications technologies (GPRS, UMTS, Wi-Fi, Bluetooth), actually quite common in mobile devices, enforces applications robustness allowing not only for horizontal but also vertical handoffs.

Next we present some prototype features and graphical user interface of NutriMe. User personal profile form is presented in Figure 6 (left picture) for user characterisation features of Definition of Personal and Nutritional Profiles Use Case. And interactive consumption registration form is also presented in Figure 6 (right picture) for Consumption Monitoring Use Case features.

![Figure 6: NutriMe personal profile form and Interactive consumption registration form.](image)

NutriMe user interface also provides friendly graphical classifications of nutritional and health ratios, such as the BMI - body mass index (Figure 6), which are subject of time evolution graphical or text based presentation and analysis.

Supporting distributed relational database technologies and strategies for NutriMe desktop and mobile prototypes are presented in section 5.

5 DISTRIBUTED DATA MODEL

Available commercial DBMS (e.g. Microsoft SQL Server) allow for flexible data management, enhancing application robustness, high availability and performance, being especially important in distributed systems involving mobile devices with non continuous communication connectivity. Replication strategies are usually adopted (and very effective) to overcome periods of non connectivity. However, data replication needs specific mechanisms for integrity enforcement among the copies spread throughout the several components of a distributed system (e.g. mobile and desktop equipment). While subject to isolated processing during non connectivity periods, data replicas in different processing devices can evolve to divergent, eventually inconsistent states. Automatic
mechanisms must be activated to rollback or roll-forward the replicas to get the overall distributed system back to a consistent state. The mechanisms adopted in our system for this purpose are based in the two-level master-slave transaction model (Liu et al., 1999).

One level is ruled by desktop workstations, the second level is ruled by the mobile devices (one or more devices). Data convergence actions take place when a communication channel is available between desktop equipment (master replicas) and mobile devices (slave replicas).

First, the master (re)executes all transactions made available by the slave, corresponding to all the actions performed by the slave during non connectivity periods. Second, the master notifies the slave the successful reconciliation transactions to be committed. Inconsistent transactions are tracked back until its root (causal graph node), and undone until the overall distributed system gets into a consistent status.

The presented two-level master-slave replication strategy seemed to us a suitable solution for the NutriMe system. It revealed optimised trade-offs according to quality of service and resource usage, in scenarios similar to ours.

6 CONCLUSIONS

This paper presents a software based system – NutriMe, which addresses ubiquitous nutritional monitoring and advising supported by several cooperating software components distributed through desktop and mobile devices, databases and applications. The system intends to introduce increasing computational and communication pervasive features, assisting in simple daily tasks with low or non intrusive reflexes (low or no perception or interaction required from users). Because aged populations are common in developed countries, and they have usually special needs, concerns of accessibility and usability have been of major importance in the design and implementation of NutriMe prototype. Individual and public interest (e.g. health, economic) of this kind of systems are worth to mention. They constituted our first interest and justified further work and research on this topic. Future work milestones include information gathering and processing in public spaces context, multimodal interfaces for impaired people, introduction of data mining techniques for customised individual user advise, extraction and analysis of general nutritional tendencies and patterns.

REFERENCES


REPLICATION OF WEB SERVICES FOR QOS GUARANTEES IN WEB SERVICE COMPOSITION

Dirk Thissen and Thomas Brambring
Department of Computer Science, RWTH Aachen University, Ahornstr. 55, 52074 Aachen, Germany
thissen@nets.rwth-aachen.de

Keywords: Web services, replication, quality of service, proxy.

Abstract: The concept of web services defines a middleware for implementing distributed applications independent of used platforms and programming languages. When developing new software systems, re-use of functionality of existing services can be done to reduce development time and costs. This process of re-use is called web service composition. But, current web service standards are not equipped to consider non-functional requirements, i.e. quality of service (QoS) aspects of a user to a composed service. Thus, capabilities of composed services cannot be guaranteed. This paper presents an approach to integrate QoS aspects into the composition of web services by using service replication. At composition time, service instances are chosen depending on the QoS requirements of a user to the whole service, and it is decided which services in the composition have to be replicated and which replication strategy to use. Replication ensures that the QoS requirements are not only considered at service selection time, but also can be granted at service runtime.

1 INTRODUCTION

A web service is some software which is seen as a service it offers, aiming at automatic machine-to-machine communication independent of the implementation of the software. A software architecture is defined to give standards for service definition (WSDL) and interaction (SOAP). Web services are loosely coupled; the search for services to communicate with is done dynamically at runtime using a service registry (UDDI). The interaction across platforms and programming languages enables easy and fast deployment of new software: complex software systems can be plugged together from existing services to a collaborating group. This is called web service composition. One vision is to achieve an automatic composition and interaction of services. But, the standards are not equipped to consider non-functional demands to a composition, i.e., QoS requirements like performance, reliability, or cost.

For acceptance by customers, a business should provide good quality of its composed services. This paper focuses on handling QoS aspects in web service composition. The service registry UDDI was enhanced to select web services for a composition with respect to the QoS requested by a user. A QoS broker was added to UDDI to manage the services’ QoS information and to calculate the best combination of services in the composition due to a user’s requirements. But, QoS aspects can be dynamic, so a QoS-oriented selection is only considering a kind of system snapshot. Depending on the time between service selection and the execution of a selected service in the composition (basing on the composition pattern and the interaction between the services), QoS information used at selection time can be outdated at service usage time. Thus, the architecture is enhanced by service replication to guarantee the QoS from selection time also at runtime. A flexible replication framework was developed to allow for as well performance-related as fault tolerance- and availability-related replication of services.

The paper is structured as follows. Chapter 2 presents a short overview about related work in the area of QoS and replication in web service composition. In chapter 3, the principle of the QoS broker is explained. Chapter 4 describes the replication architecture and gives an overview about the current implementation status. Finally, chapter 5 concludes the paper and gives an outlook on the ongoing work.

2 WEB SERVICES AND QOS

Several approaches for the composition of web services exist. A prominent example is the Business
Process Execution Language (BPEL) for Web Services (OASIS, 2007). But these approaches are not considering quality of service in the composition process. Though there is a lot of research in web services and composition, not much is related to QoS. The existing work mostly refers only to a part of the whole problem. (Liu et al, 2004) for example present a framework to publish up-to-date QoS information for web services, but the success of this mechanism depends on feedback of users about the quality of the services they consume. (Zeng et al, 2004) present a method to select services that fit to a user’s interest (expressed as QoS parameters). Local optimization and global planning are combined to find the best set of services for a composition. But, in case of highly dynamic QoS parameters, the global planning approach might take more time for re-calculation than the execution of the service would need. Thus the approach itself can violate the QoS. (Jaeger et al, 2004) propose a mechanism which could be more efficient by using an aggregation scheme for QoS aspects. The approach sounds well but was not implemented nor tested by the authors. We used this approach as basis for the implementation of an own solution to consider QoS aspects in composition (Thißen and Wesnarat, 2006) which is explained more detailed in chapter 3.

All these approaches have the same weakness: services for the composition are chosen some time before execution. If QoS parameters change, during service execution the QoS demands of a user nevertheless can be violated. Replication is a possible solution to deal with dynamic QoS parameters on performance, high availability, and fault tolerance/reliability. Instead of running a single instance of a service, several copies are used. Replication defines methods for keeping consistent all copies (called replicas). The complexity is hidden from the user of a service by a frontend which acts as the service from the user’s view. A lot of replication algorithms are given. In active replication, all replicas act in the same way. The frontend uses group communication to distribute a request to all replicas. It decides how to deal with responses of the replicas, depending on the QoS aspect which should be considered. To ensure service available or to decrease the response time of a service (performance), the frontend returns the first response to the user. To improve fault tolerance, it compares and combines all responses. A different approach is passive replication. One replica is a primary, and the frontend only communicates with this replica. The primary forwards the requests to all other replicas (backups) to keep them consistent. If the primary fails, a backup can take over its role, which improves fault tolerance and availability.

There are lot of other replication algorithms, and also approaches exist to implement them within a web service architecture. E.g., (Ye and Shen, 2005) discuss active replication for web services. But, the focus is only on reliability of web services, and only active replication is implemented. The same holds for (Chan et al, 2007): it is focussed on reliability. WS-Replication (Salas et al, 2006) also uses active replication to achieve high availability, and WS-multicast is used for communication between the replicas. WS-multicast is SOAP-based and maybe causes a high overhead. (Osrael et al, 2007) is a more flexible approach, implementing passive replication and designing an open system for later addition of other replication strategies. Consistency can be weakened in this approach to reduce the performance overhead caused by update propagation. But, till now only a variant of passive replication is realized, and the focus is on fault tolerance.

Concluding, the replication approaches either focus on only one replication strategy, use multicast on SOAP level which decreases performance, or only consider a certain QoS aspect, e.g. availability. Thus we designed an own replication framework for integration with composition, which offers more flexibility, see chapter 4.

3 QOS IN SERVICE SELECTION

For composing web services under QoS constraints, we followed the approach presented in (Jaeger et al, 2004): a workflow pattern is given, showing the relations between services. Aggregation rules are used to combine quality measures assigned with single services to come to an overall rating of sets of services. We identified relevant QoS information and basic composition patterns (SEQUENCE, AND, OR, XOR, and LOOP) from which the whole workflow pattern can be formed. Next, we defined corresponding aggregation rules and a selection mechanism to choose the best service candidates. Given the workflow pattern for a composed service, aggregation of QoS parameters is done by collapsing the whole composition graph step-wisely into a single node, starting with the innermost composition pattern. By aggregating the properties recursively, only one node is left in the final state. A set of formulas was defined to model the aggregation of the QoS parameters performance, cost, reliability, and availability. This mechanism enables us to check the resulting QoS of a set of services. Because
we need to find a set of services to be executed, each possible combination of service candidates for the current composition pattern is evaluated, and the best ones regarding the user’s demands are selected. Multiple criteria decision making and weighting are used to combine a service set’s aggregations for different QoS parameters into one value, a quality score. For a composition pattern, the set of service candidates with highest quality score is selected, and it is done aggregation of the next innermost composition pattern till a single node (the composed service) remains with assigned QoS values.

For implementation of this approach we have designed a prototype which enhances the general web services architecture. *Apache tomcat* was used as web container for the provided web services, *Apache Axis* services as SOAP implementation. *jUDDI* was chosen as UDDI registry, for executing composed services the *Oracle BPEL Process Manager* was used. It provides a service orchestrator which can be assigned a workflow pattern and a set of basic services; it then manages the execution of the services due to the pattern. For integrating QoS consideration as described before, we have implemented some more components, see figure 1.

The central component is the QoS broker which implements the QoS aggregation rules. It involves a BPEL registry and a QoS registry. Service providers as usual register their services with UDDI (step A in figure 1). To publish QoS information, a monitor is assigned each service, registering with the QoS registry when a service is put into UDDI (step B). When a composed service is deployed, the workflow pattern is stored in the BPEL registry (step C).

When a service requestor searches for a web service, it contacts the QoS broker (step 1). It does not need to know if a service is a composed one or not; the broker uses the BPEL registry to search for a composition pattern. If one is found, the broker asks UDDI for available candidates to all services in the pattern (step 2). Having retrieved a list of all available candidates (step 3), the broker connects to the QoS registry to get the QoS values for these services. By stepwise aggregation of the values according to the pattern from the BPEL registry and by selection of the best fitting candidates, a set of basic services is chosen (step 4). The requestor gets back a reference to an orchestrator for using the service (step 6/9). The orchestrator manages the service execution (step 7/8). After execution, it gives feedback to the broker. In other requests to the same composed service, the broker can make use of it.

Not included in figure 1 is the use of the QoS monitors. Getting the QoS information for aspects like cost is no problem: the values are constant for a longer period of time and can be filled in by the service provider at service setup. But most aspects, are dynamic, e.g. like performance. Thus a monitor is assigned each service to record its behaviour, to compute floating averages, and to forward this information to the QoS registry. To avoid that service providers have to modify their services, the monitors are independent components. They get the needed information from so called valves placed on Tomcat engine level. Here, e.g. timestamps can be used to get statistics about the queuing time of a request.

Nevertheless, the QoS broker cannot guarantee the QoS from selection time to be constant at runtime, thus we had to enhance this architecture by a mechanism which allows for some control at execution time of the services.
4 REPLICATION FOR QOS GUARANTEES AT RUNTIME

The architecture presented in chapter 3 is only able to consider user demands at selection time. Our next step was to enhance the architecture by capabilities of replication, to control the selected QoS at runtime. Because of the disadvantages of existing approaches, we designed an own replication architecture considering the following goals:

- **Allow for flexible choice of replication algorithm at runtime.** We want to use replication for guarantees on several QoS aspects, thus we need different replication strategies supporting performance, availability, and fault tolerance in one approach.
- **Open architecture which can easily be enhanced with new replication algorithms.** For the beginning, we only considered the most prominent algorithms: active and passive replication.
- **Decide at composition time which services have to be replicated to fulfill a requestor’s demands.** E.g., the use of several replicas may improve the reliability, but may contradict the cost of service usage if one has to pay for each extra replica. Thus, in the selection process a tradeoff is necessary between gain and costs of using replication, including the number of replicas to use.
- **Transparentsly use group communication and avoid communication overhead by using SOAP.** Otherwise, replication could contradict the QoS.
- **Automatically generate request and result classes for web services from WSDL files.** Reduce the costs and time for integrating the mechanisms into each application newly.

We designed our replication architecture oriented at these goals and allowing for easy integration with our QoS-based selection of service candidates in a composed service (chapter 3). In the following, the components of the architecture and their interaction are described in more detail.

The **QoS broker** remains the central component of the architecture. It is enhanced by enabling the selection of a replication strategy as well as a set of suitable replicas for a service. For simplicity, we started with the consideration of simple services within the replication process, but oriented at the composition architecture for easy integration.

The interaction of the QoS broker with service requestor and the replicas of a single service is shown in figure 2. The replicas are all registering with UDDI as usual (step (a) in figure 2). The replicas additionally register with the QoS broker resp. the assigned QoS registry via their monitors (b) as described in chapter 3. The services’ monitors do not need to know if they are belonging to a replicated service or to a simple one, they have to submit the same information as before. The monitors regularly measure the QoS values of their replicas, calculate advanced information like floating averages, and deliver the resulting values to the QoS broker.

![Figure 2: Replication enhancement.](image)

If a service requestor contacts the QoS broker to ask for a service (1), the broker interacts with UDDI to find all replicas to the requested service (2 + 3). Based on the requested QoS, the broker now can select a subset of fitting replicas which seems to be sufficient to fulfill the requestor’s demands. Simultaneously, it can decide on the best replication strategy regarding the requested QoS. If e.g. high performance is needed primarily, active replication is chosen to reduce response times. If availability has priority, passive replication is more appropriate to reduce the communication overhead. Based on the known availability probability of the service, also the number of replicas could be determined. Currently, active and passive replication are implemented in our prototype, and only a few rules are implemented, which strategy to use in which cases.

The service requestor gets back a reference for its service (4) and can use it (5 + 6). The detailed information transmitted in these steps depend on the replication strategy chosen by the QoS broker since the service requestor maybe has to contact a single service or maybe a service group.

If the broker chooses passive replication, the requestor only communicates with a single replica. In contrast to common passive replication there is no fixed primary replica which all the time is contacted. Instead, the QoS broker chooses the actually best replica due to the requestor’s demands and returns a reference to this replica to the requestor. The other
replicas only serve as backups and are invisible to the requestor. The dynamic primary selection allows for a better average QoS level in terms of performance because it enables a kind of load balancing between all available replicas. But, one has to keep in mind that using different primaries for different requests can cause consistency violations. Thus it depends on the service itself if the weakening of the consistency is useful.

On the other hand, in active replication the requestor has to communicate with all replicas simultaneously. Thus he has to communicate with a group of services instead with a single service.

To hide the different usage for the replication schemas, proxies are used. They encapsulate the functionality of communication with replicated services. Only a single interface is offered to the requestor. Independent if passive or active replication is used, the requestor gets back a reference to the used proxy instead of a reference to a concrete service (in step 4 of figure 2) – the proxy itself seems to be the service for the requestor. This schema intentionally is designed similar to the usage of composed services via an orchestrator as described in chapter 3, to merge the functionalities of orchestrator and proxy. The only difference for the requestor is that the QoS broker not only sends back a reference to a service (the proxy), but also some additional configuration parameters the requestor has to use in its request to enforce a certain replication process (which was chosen by the broker).

![Figure 3: Client proxy.](image)

The original request of the user is passed to the proxy only as one parameter request. The proxy is able to process this request by using the correspondingly assigned replicas. In which way to use the replicas is defined by the other parameters of the user’s call. The proxy implements functions requestActively and requestPassively. Depending on which replication mechanism is chosen the client has to call the corresponding function. The client gets this information from the QoS broker as part of the configuration information. The second parameter of the request tells the proxy if the request is read-only or not. In case of read-only, consistency is relaxed, which can improve the performance of a request. This parameter is followed by an information if the first response has to be forwarded to the client (e.g. for performance or availability aspects), or if the proxy has to wait for all responses and to combines them in some way to achieve fault tolerance. The last parameter is a timeout. It defines how long the proxy has to wait for responses before combining the received results (or before sending an error message back to the requestor).

The proxy now can inform the group communication component about the needed communication mechanism (2) and the request correspondingly is passed only to a single service or to a group of services (3). The results which are coming back from the replicas (4) are passed on to a message handler (5) which can treat the responses in different ways as described above. If passive replication was used, the proxy immediately uses a callback function to deliver the result to the requestor (6). In case of active replication, it can forward the first response to the client, or collect all requests coming in before a timeout and form a consensus out of them before passing only a single response to the requestor.

Also on server side a proxy is needed to coordinate all replicas corresponding to the chosen replication strategy, see figure 4. The request comes in over the group communication mechanism (step 1 in figure 4) and is forwarded to the message handler (2). The message handler in the background interacts with the QoS monitor (a) to allow for statistics about the number of requests per second, response times, etc which is part of the QoS parameters collected by the monitor. Because in active replication consistency requires a sorted execution of requests in its request (step 1 in figure 3). Such a request may look as follows:

```
Proxy.requestActively(request, READ_ONLY, GET_FIRST, 2000);
```

The original request of the user is passed to the proxy only as one parameter request. The proxy is able to process this request by using the correspondingly assigned replicas. In which way to use the replicas is defined by the other parameters of the user’s call. The proxy implements functions requestActively and requestPassively. Depending on which replication mechanism is chosen the client has to call the corresponding function. The client gets this information from the QoS broker as part of the configuration information. The second parameter of the request tells the proxy if the request is read-only or not. In case of read-only, consistency is relaxed, which can improve the performance of a request. This parameter is followed by an information if the first response has to be forwarded to the client (e.g. for performance or availability aspects), or if the proxy has to wait for all responses and to combines them in some way to achieve fault tolerance. The last parameter is a timeout. It defines how long the proxy has to wait for responses before combining the received results (or before sending an error message back to the requestor).

The proxy now can inform the group communication component about the needed communication mechanism (2) and the request correspondingly is passed only to a single service or to a group of services (3). The results which are coming back from the replicas (4) are passed on to a message handler (5) which can treat the responses in different ways as described above. If passive replication was used, the proxy immediately uses a callback function to deliver the result to the requestor (6). In case of active replication, it can forward the first response to the client, or collect all requests coming in before a timeout and form a consensus out of them before passing only a single response to the requestor.

Also on server side a proxy is needed to coordinate all replicas corresponding to the chosen replication strategy, see figure 4. The request comes in over the group communication mechanism (step 1 in figure 4) and is forwarded to the message handler (2). The message handler in the background interacts with the QoS monitor (a) to allow for statistics about the number of requests per second, response times, etc which is part of the QoS parameters collected by the monitor. Because in active replication consistency requires a sorted execution of requests in its request (step 1 in figure 3). Such a request may look as follows:

```
Proxy.requestActively(request, READ_ONLY, GET_FIRST, 2000);
```

The original request of the user is passed to the proxy only as one parameter request. The proxy is able to process this request by using the correspondingly assigned replicas. In which way to use the replicas is defined by the other parameters of the user’s call. The proxy implements functions requestActively and requestPassively. Depending on which replication mechanism is chosen the client has to call the corresponding function. The client gets this information from the QoS broker as part of the configuration information. The second parameter of the request tells the proxy if the request is read-only or not. In case of read-only, consistency is relaxed, which can improve the performance of a request. This parameter is followed by an information if the first response has to be forwarded to the client (e.g. for performance or availability aspects), or if the proxy has to wait for all responses and to combines them in some way to achieve fault tolerance. The last parameter is a timeout. It defines how long the proxy has to wait for responses before combining the received results (or before sending an error message back to the requestor).

The proxy now can inform the group communication component about the needed communication mechanism (2) and the request correspondingly is passed only to a single service or to a group of services (3). The results which are coming back from the replicas (4) are passed on to a message handler (5) which can treat the responses in different ways as described above. If passive replication was used, the proxy immediately uses a callback function to deliver the result to the requestor (6). In case of active replication, it can forward the first response to the client, or collect all requests coming in before a timeout and form a consensus out of them before passing only a single response to the requestor. 
from different clients on all replicas, the holdback queue (3) delays all requests till they can be executed without violating consistency to other replicas. To do so, requests have to be sorted the same way for all replicas. For this purpose Lamport Timestamps are used. The requests are sorted into a delivery queue (4) which simply implements a FIFO strategy and executes one request to the service after the other. The requests can be passed on to the web service by using a callback function (5 + 6). The delivery queue also gets back the response (7 + 8) and initiates the transmission of this response back to the requestor (9). Again, the group communication mechanism takes over the transmission of the result to the requestor (and to the backups, in case of passive replication).

Using client and server proxy, the whole replication is transparent for users and services. When new replication strategies are implemented, only the proxies have to be enhanced.

5 CONCLUSIONS

Currently, the architecture as described in chapter 4 is finished, and experiments are performed to evaluate the behaviour of the replication framework. On one hand the experiments should validate the correctness of the implementation. On the other hand (and more important for the ongoing work) the evaluations also should help in comparing gains and costs of the replication strategies. These comparisons are necessary for fine tuning of the decision rules inside the QoS broker: for which combination of requested parameters which strategy should by used, and with how many replicas. In parallel, implementation has started to integrate the replication enhancement into composed services. This task is easy because the architecture of the replication system was oriented at the existing composition architecture (integration of orchestrator with client proxy, implementation of server proxy as valves like the monitors). Afterwards, the gain of using replication in the composition again has to be evaluated by a number of experiments.

Replication only is one way to improve the quality of a service. After finishing our current work, beside integrating more replication strategies we want to examine if instead executing the same service several times, also equivalent services of different providers could be used. Also, we plan to enhance the functionality of the proxies by other strategies, e.g. load balancing as a mechanism with weaker guarantees as replication, but on the other hand cheaper if services – and quality guarantees – have to be paid for.

REFERENCES


AN E-VORTAL FOR THE PORTUGUESE BAKING INDUSTRY

Requirements Model

Jorge Gouveia, Paula Oliveira
Department of Engineering, University of trás-os-Montes e Alto Douro, Quinta de Prados, Vila Real, Portugal
jgouveia@utad.pt, poliveir@utad.pt

João Varajão
Department of Engineering, University of trás-os-Montes e Alto Douro, Quinta de Prados, Vila Real, Portugal
jvarajao@utad.pt

Keywords: e-vortal, Internet, Baking industry.

Abstract: Nowadays, Internet is fundamental to the development of the competitive capability of practically any industry. Web portals, mainly the vertical portals, have in this context a major role, because they are basic instruments for the dissemination and search of the information. Despite this fact, it is apparent that no such Web solution is available in Portugal for the baking industry, in spite of the fact that this is one of the major industrial sectors in the national context. Believing that the lack of a vertical portal for the baking sector is a deterrent to its growth, in this paper we present the results from a study developed to identify the main requirements for a system of this kind.

1 INTRODUCTION

The creation of a web portal for the baking industry is, for some years now, a reality in several European countries and other world markets. The reasons invoked for its creation are most diverse. One of them is the easiness and comfort with that the portals supply the access to the information, in a versatile, customizable and personalized way, based on the preferences of the users. Another frequent described reason is the globalization. Globalization causes an increase of the competition due to bigger integration of the markets (Gouveia et al., 2007a).

In spite of the fact of the baking industry is one of the major industries in Portugal, the Portuguese scenario is somewhat different: up until now there is no such portal (Gouveia et al., 2008).

The majority of Portuguese food and agriculture companies have an incipient use of information technologies when compared to their European counterparts (SPI, 2001). Though the trend is to grow, the companies’ use of the Internet to buy and sell goods and services is relatively negligible (INE, 2001, INE, 2003, INE, 2004, INE, 2005).

The sector’s lag in this domain, together with the lack of a dedicated e-vortal, is one of the main motivations to this work: to contribute to the conception of an Internet solution (vertical portal), that gathers information, services and other resources useful to the various intervening parts, by identifying its main requirements.

Technological innovations may contribute in numerous ways to a company’s competitive advantage, by enabling improvements in satisfaction of demand, cost reduction or quality increase. So, this kind of portal can act as a basis for new economies and be a booster in lifting old economies to the new market realities (Zirpins et al., 2001).

Thus, we propose the creation of an e-vortal for the Portuguese Baking Industry: a portal that provides to all the community an access point to the information and resources, in a way to potentially increase the business between the suppliers and the baking industry companies.

In this paper we propose a set of main requirements for the development of vertical portals in baking industry. In section 2 we introduce some fundamental concepts on Internet web portals and vertical portals and make a brief presentation of the baking industry in Portugal. Section 3 presents several requirements that should be considered in the design of new vertical portals for the baking industry. The paper ends with some final remarks in section 4.
2 CONTEXT

In the past few years, the Internet and, particularly the World Wide Web (WWW) has expanded continuously, in terms of the technology used and in terms of dimension, becoming an essential way for business relationships. From this development, several instruments of research and data organization appeared, in which the Web portals have a particular relevance (Gouveia et al., 2007).

2.1 Web Portals

The concept of Internet “portal” is relatively recent and it is not consensual yet. One can briefly define a portal as an integrating point of access to relevant information, a convergence point for different users, with a wide range of information.

The idea of Internet Portal is comprehensive and may include everything from a simple online catalogue to a complex intranet solution. However, a common characteristic is its role as a starting point, offering a doorway into using web services (Zirpins et al., 2001).

The fast growth of the WWW and the recognized importance of Portal Web, led in few years, to the incoming of countless portals, many times with different characteristics among themselves. Consequently, several authors proposed different definitions, many times for the same object of interest (Gouveia et al., 2007).

One of the first classification of portals was presented in 2002 (Strauss, 2002), dividing portals into two large groups according to the depth of their content: horizontal portals and vertical portals. Horizontal portals are public websites whose goal is to serve its users with a broad set of services and resources in an effort to convince them to make the website their homepage. On the other hand, vertical portals provide information, application and other resources targeted to a specific community or interest group.

In 2003, Clarke and Flaherty added two new dimensions to this classification: not only the depth of the portal’s content (vertical/horizontal), but also its mission (transaction/information) and its target (public/private). Each dimension should be viewed as a continuum, with all portals expressing varying degrees of each element (Clarke and Flaherty, 2003).

Focusing our attention on the “depth of content” dimension, vertical portals, or “vortals”, offer contents and services targeted to a specific domain or community: professional classes, people from certain locations or with other common interests (Zirpins et al., 2001).

Given the increasing difficulty of efficiently searching the web using keywords, a new trend in the design of website for specific users appeared: a verticalization phenomenon, i.e., the creation of portal-like internet websites specially tailored for a specific subject or area of interest, which allows for the reduction of portal size without omitting relevant content.

As opposed to general-purpose search engines, vertical portals have search tools with adequate strategies and terminology, focused on the target market (Medeiros et al., 2000).

A vortal, an abbreviation of “vertical portal”, is a vertical industry, market or specific group, portal on the Internet. Vortal refers to a website that aggregates varied content and services of interest to a particular industry and makes it available to industry members. While a portal, such as Yahoo! or AOL, attracts a large number of “netizens” (Internet citizens), offering a wide range of contents and links to other sites, vortals are narrower in focus and address a specific industry, theme, or interest. The audience or participants may be narrower, but they have a higher interest and are highly targeted (Vortalbuilding, 2005b).

The concept of a vertical portal is based upon the premise of linking customers and vendors together within a focused environment. This environment provides information, services and other resources, which encourage members to remain within the boundaries of the vertical portal. Vortals are also seen as business-to-business communities or business-to-consumer communities (Vortalbuilding, 2005b).

A well-designed and well-developed vertical portal can create a snowball effect. Users visit it because it has quality information, advertisements, discussion forums, products, contents, friends, etc. A higher number of website users, imply more advertisements, products and participation, which in turn will attract more users to it. Once the users’ loyalty is ensured, they will keep visiting and using the portal on a regular basis.

Companies and individuals who share the same interests may gather to interact, collaborate and transact on a digital market. In Portugal, such need clearly exist for certain markets, as one can infer from the recent appearing of industry specific portals, like the construction industry.
2.2 The Portuguese Baking Industry

According to the Portuguese classification of economic activities (CAE), the Baking and Pastry Industry is part of group 158 – Manufacturing of other alimentary products. Group 158 and others make the Food, Beverages and Tobacco Industry division.

In terms of gross sales, group 158 is the second largest in its division, just behind group 159 (Beverages Industry). The Baking and Pastry Industry alone (CAE's class 1581) contributed, in 2001, with 49% of group 158's gross sales (INE, 2002a).

The relative weight of class 1581 has been growing since 1998. That year, according to INE (INE, 1999), this economical sector represented just 28% of group 158’s total gross sales.

If gross sales-wise group 158 is relevant in the food industry context, in terms of employment it is even the more so. Group 158 – Manufacturing of other alimentary products – is by far the most important, representing approximately 50.000 jobs in the year 2000.

Thus, the importance of the baking industry in the context of the Portuguese food industry is evident.

2.3 Portuguese Baking Industry

Vertical portals design and development is, for some years now, a concern and a reality in several countries (Maltz, 2005).

In Portugal the reality is quite different from other countries. The e-vortal concept is not yet well established, as we can see by the lack of implemented e-vortals focused on the Portuguese Industry. If we focus our attention on the specific sector of baking industry, then the situation is extreme. After an exhaustive Internet search we concluded that currently there is no an e-vortal for the Portuguese baking and pastry industry (Gouveia, 2006).

There are, certainly, a few sites that belong to industrial associations of the sector which aren’t more than an institutional presence in the web. We can, as an example, refer to the site of the FIPA – Federação das Indústrias Portuguesas Agroalimentares (Food-Agricultural Federation of Portuguese Industry), ACIP-Associação do Comércio e da Indústria de Panificação, Pastelaria e Similares (Baking Trade and Industry Association) (Gouveia et al., 2007a).

The FIPA was constituted in 1987 with the aim to represent and defend the interests of the Portuguese Food-Agricultural Industry nationwide and in the European Union. This site is dedicated to every Portuguese Food-Agricultural industry, as well as those that work directly with them. In spite of not being an association of the Bread Makers sector, it has yet some connection with it. This site can be classified as an informative portal. In it one may find, essentially, relevant information for to the Portuguese Food-Agricultural Industry. Another case is the official site of the Trade Association (ACIP). This association presents itself as the largest Portuguese association of the Bread Making and Bakery sector. This site can be classified as an institutional one, once it gathers information about the ACIP as well as news and data of its activities. So far, this site isn’t providing any further service to the user (Gouveia et al., 2007a).

3 AN E-VORTAL FOR THE BAKING INDUSTRY

A well-conceived e-portal may create a snowball effect: having good contents, forums of discussion, products and advertising, people will visit the portal. As many people visit the portal, the larger will be the number of advertisers, the participations in the discussion forums, the production of contents and the possibilities of cooperation and data sharing, therefore contributing to the development of industry.

The bakery sector is one of the most representative sectors in the Portuguese industry as a whole (INE, 2002a).

The lack of Web vertical portal hinders the chances of a bigger development for the baking and pastry industry. So, we propose a new e-vortal for the Portuguese baking industry.

When one considers the creation of an industry-specific portal for a sector such as the baking industry, the processes involving the identification, analysis, negotiation, description, validation and requirements management assume a paramount importance. These processes are decisive in the system development and assume a great relevance as a factor of success in its construction (Gomes and Soares, 2004).

In this section first we discuss the fundamental concepts that should be considered in the development of an e-vortal for the Portuguese baking industry. Next, we present the requirements that a system of this kind should support.
3.1 Fundamental Concepts

An e-vortal solution for the industrial sector should be based on five fundamental concepts, as depicted in Figure 1 (Gouveia et al., 2008):
- Community;
- Collaboration;
- Interactivity;
- Contents;
- E-Commerce.

3.1.1 Community

The main aim of an industry vertical portal is to encourage companies and individuals with interests in a specific industry to interact, collaborate and do business within a digital market. This virtual community will allow actors with a common interest to meet, to share ideas and information, and to know each other better, so they can strengthen their relationship.

3.1.2 Collaboration

It is important to distinguish the words “collaboration” and “cooperation”. Usually we don’t make a distinction between them (Dillenbourg and Schneider, 1995). However, there is a difference in the way the activity is performed by the ensemble. Cooperation consists in dividing tasks among the participants, each person or organization being individually responsible by part of the problem solution. Collaboration is characterised by the mutual contract of the participants, who work in a coordinated effort to solve the problem all together (Dillenbourg and Schneider, 1995). Sharing experiences, searching for new solutions and products in partnership, are examples of possible collaboration in the baking sector of industry.

3.1.3 Interactivity

Internet websites and portals in particular, offer lots of communication opportunities in both directions: they offer interactivity. Stuart Brand defines it as a “Mutual and simultaneous activity on the part of both participants, usually working towards some goal, but not necessarily” (Brand, 1987). With an e-vortal, the user should be able to interact with the system and find the information she or he needs.

3.1.4 Contents

The content of a vertical portal is very important. The existence of e-vortals is in part due to a huge need to reduce the size of portals, making search easier, while no omitting relevant contents. The search tools can (and should be) improved and they must provide the specific terminology and strategies of searching for a sector or subject (Vortalbuilding, 2005a). The users’ fidelity is one of the greatest challenges that an e-vortal brings up. The quantity and quality of its contents is an important item to achieve that fidelity. If we keep the contents always up-to-date with a high level of quality, existing users will remain faithful to the portal and more new users will visit it, which in turn will make the portal even more attractive to the participation of different agents – partners, announcer and collaborators. The increase of participation will allow a higher degree of information sharing and higher possibility of collaboration among partners.

3.1.5 e-Commerce

e-Commerce is the purchasing or sale of goods and/or services through electronic networks such as the Internet. A platform which will make relationships and business transactions easier between partners should be provided in an e-vortal for the baking industry. This will improve a number of aspects of a business, both for sellers and buyers.

Being one of the most complex sides of a vertical industrial portal, we can consider two phases of its development. In a first phase it can promote the relationships and transactions among business partners. So, an e-vortal will allow several aspects of the business process, for the one who sells, as well as for the one who buys. The promotion of new products, the small effort demanded by the support of catalogues, the quick answer to the customers and
the possible increase of sales due to the increase of the market base, are some of the advantages foreseen for the sellers.

The fastest answer for the search of the best solutions, the quick access to new suppliers, and consequently to better prices are, among others, some of the advantages for the buyer.

In a first stage of the portal for the Baking Industry, one can say that it will improve the communication among several agents. The partners will be able to decide more easily as they share a common platform.

For a second stage, it will be reserved the b2b transactional platform. The implementation of automatic processes of proposal requests, the submission of proposals and the closure of deals are the aspects taken cared by the platform.

3.2 Requirements

The design of a new Baking Industry e-Vortal should take in account all the concepts referred in previous section. In this section we identify and describe some of the function/components such e-vortal should support.

A study has been made with the purpose of investigating different aspects of vertical Internet portals. It was an interpretative and exploratory study and consisted, in a first phase, on the development of a conceptual framework and, in a second phase, on the identification and analysis of currently existing vertical portals, trying to understand their objectives, characteristics and functionalities. The needs of the baking industry were particularly studied.

Thus, four dedicated e-vortals to the bakery industry had been analyzed: Bakery-Net (www.bakery-net.com); Bakery Online Market Place (www.bakeryonline.com); Federation of Bakers (www.bakersfederation.org.uk); e Portal da Padaria (www.portaldapadaria.com.br).

This research made possible the identification of various critical elements that were systemized in the form of a set of requirements that are presented in table 1.

The display of the main characteristics/functions for an e-vortal solution for the bakery sector comes from a conjunction of characteristics found in other portals, as in portal bakery-net.com, with other completely new, as a result of the analysis of the sector needs.

Table 1: E-vortal main requirements.

<table>
<thead>
<tr>
<th>Function</th>
<th>Brief description</th>
</tr>
</thead>
<tbody>
<tr>
<td>Bakers’ Guide</td>
<td>This is an area where all companies registered should be listed. Some groups can be: Equipment vendors, Ingredients, Maintenance &amp; Sanitation, Packaging, Shipping, Services and others</td>
</tr>
<tr>
<td>Documents</td>
<td>Publishing and browsing documents and papers related with the bakery sector</td>
</tr>
<tr>
<td>Legislation</td>
<td>Browsing important legislation for the baking industry sector</td>
</tr>
<tr>
<td>Useful information</td>
<td>Useful information like weather and stock market information</td>
</tr>
<tr>
<td>News</td>
<td>News about the industry</td>
</tr>
<tr>
<td>Event Scheduler</td>
<td>Browsing and searching on event scheduler</td>
</tr>
<tr>
<td>Training Courses</td>
<td>Here it should be possible to browse all the training courses available. It should be also possible to do the registration in a course available</td>
</tr>
<tr>
<td>Recipes</td>
<td>Recipes browsing and sharing</td>
</tr>
<tr>
<td>Classified ads</td>
<td>Browsing and creation of small classified ads</td>
</tr>
<tr>
<td>On-line Catalogues</td>
<td>Products browsing and online catalogue</td>
</tr>
<tr>
<td>B2B platform</td>
<td>This platform act as a promoter of the relationships and transactions between partners. Thus, several aspects of the business processes will be made easier (for buyers and sellers). Requests for additional information about equipment, products or services, delivery or terms of payment, are some examples.</td>
</tr>
</tbody>
</table>

As already referred, web portals can be classified according to three dimensions (Clarke and Flaherty, 2003): ) the portal purpose, contents and the level of access provided. The portal one proposes presents these characteristics in several levels, as in Figure 2.
The e-vortal main requirements presented on table 1 will be detailed and classified according to these dimensions in the following sub-sections.

3.2.1 Portal Purpose

According to their purposes, the portals can be divided in transactional and informational. The solution presented cannot be classified, exclusively, as an informational e-vortal. It will contain characteristics mainly of an informational portal with updated news of the sector, newsletter, events and training courses, updated legislation, among others. It will also have transactional portal characteristics. This portal will allow customers gathering data of products (characteristics, prices, conditions of delivery, availability, etc.), comparing prices and, mainly, promote business transactions. In Table 2 one can see the main characteristics of the proposed e-vortal according to its purpose.

<table>
<thead>
<tr>
<th>Transactional</th>
<th>Informational</th>
</tr>
</thead>
<tbody>
<tr>
<td>Online Catalogues</td>
<td>News</td>
</tr>
<tr>
<td>Products and prices comparison</td>
<td>Newsletter</td>
</tr>
<tr>
<td>B2B platform</td>
<td>Legislation</td>
</tr>
<tr>
<td>Training courses - browsing and registration</td>
<td>Event scheduler</td>
</tr>
<tr>
<td></td>
<td>Training courses Agenda</td>
</tr>
<tr>
<td></td>
<td>Stock Market information</td>
</tr>
<tr>
<td></td>
<td>Recipes</td>
</tr>
<tr>
<td></td>
<td>Articles and other documents</td>
</tr>
<tr>
<td></td>
<td>Classified Ads</td>
</tr>
<tr>
<td></td>
<td>Baker' Guide</td>
</tr>
</tbody>
</table>

3.2.2 Content

The horizontal portals are those whose contents are wide. This kind of portals don’t give access to specific community services contents, but on the contrary, they dispose information, applications and other resources, through multiple category of users, representing every member of a community.

The vertical portals can display contents and services directed to a domain or specific community. They can be centered in specific professional communities, people from certain places or communities with common interests.

This portal is clearly a vertical one, due to the particular content and the fact that it is dedicated to a well defined community.

The classification of the proposed solution is, in this dimension, simpler. The bakery portal is mainly a vertical one, because it focused content and services, and the fact that it is dedicated to a specific community with a common interest – The Bread and Bakery Industry. In Table 3 one can see the main characteristics of the proposed e-vortal, classified according to the contents.

<table>
<thead>
<tr>
<th>Vertical</th>
<th>Horizontal</th>
</tr>
</thead>
<tbody>
<tr>
<td>News</td>
<td>Weather information</td>
</tr>
<tr>
<td>Newsletter</td>
<td>Stock market information</td>
</tr>
<tr>
<td>Legislation</td>
<td></td>
</tr>
<tr>
<td>Event Scheduler</td>
<td>Courses Agenda</td>
</tr>
<tr>
<td>Online Catalogues</td>
<td></td>
</tr>
<tr>
<td>Products and prices comparison</td>
<td>B2B platform</td>
</tr>
<tr>
<td>Recipes</td>
<td></td>
</tr>
<tr>
<td>Classified Ads</td>
<td>Articles and other documents</td>
</tr>
<tr>
<td>Baker’ Guide</td>
<td>Training courses - browsing and registration</td>
</tr>
</tbody>
</table>

3.2.3 Level of Access

A public portal is a portal of which information and/or services are available for any Internet user and have no access restrictions. Opposite to these, the Private Portals have restricted access to a group of users.

The bakery e-vortal will give free access of some contents for every Internet users. As an example there will be the news, legislation and scheduled events, for example. The Table 4 presents the e-vortal characteristics according to its users.

<table>
<thead>
<tr>
<th>Public</th>
<th>Private</th>
</tr>
</thead>
<tbody>
<tr>
<td>News</td>
<td>Registered users</td>
</tr>
<tr>
<td>Legislation</td>
<td>Newsletter</td>
</tr>
<tr>
<td>Event Scheduler</td>
<td>Recipes</td>
</tr>
<tr>
<td>Stock market information</td>
<td>Browse Articles and other documents</td>
</tr>
<tr>
<td>Weather information</td>
<td></td>
</tr>
<tr>
<td>Classified Ads</td>
<td>Partners</td>
</tr>
<tr>
<td>Training Courses browsing</td>
<td>Browse and maintenance of online catalogues</td>
</tr>
<tr>
<td>Baker’ Guide – List of companies</td>
<td>Products and prices comparison</td>
</tr>
<tr>
<td></td>
<td>B2B platform (proposals/additional information)</td>
</tr>
<tr>
<td></td>
<td>Training courses Agenda</td>
</tr>
<tr>
<td></td>
<td>Training Courses – Registration</td>
</tr>
<tr>
<td></td>
<td>Baker’ Guide – companies browsing</td>
</tr>
</tbody>
</table>
Some of the contents will be of restricted access to registered users. This restriction will become it two distinct levels: registered user and partner. The distinction between these two types of users is the following one: a registered user will have access the preferential content, such as Recipes and Newsletters; a partner, for its side, will not only have access to the previously described preferential content, but also to all the services related with the e-vortal b2b functionalities. We can highlight: on-line catalogue browsing and maintenance, products and prices comparison, buy or sell on-line, training courses, agenda and registration, among others.

3.3 Discussion

The expected impact after the implementation, in Portugal, of an e-vortal for the baking industry can be analyzed according two complementary aspects. The first aspect is about the competition, it will allow, among others things, a bigger integration of the markets, the multiplication of commercial exchanges and the increase of agreements and partnerships between companies. It will help the various actors of the baking industry to create new commercial relations, as well as fortifying the existing relations. In the “new economy”, the information, the knowledge and know-making becomes raw material, and the possibility of being able to access it, in useful time, will be one of the main strategically weapons of the competitiveness. The companies of the baking sector will have at it disposal a tool that will allow them to face these new requirements. Thus, we contribute for the satisfaction of the necessity of modernization of the sector. One another aspect is the resultant advantage of the business growth, as much for who sell as for who purchase. Reduction of administrative costs, increase of productivity and modernization of the market are some of the testimonials given to the press by company CEOs and company owners when they are mentioned to the impact of the adoption of an e-vortal tool in its company. One of the reasons pointed in practically all the testimonials is the increase of the business opportunities. The chance to negotiate with new companies is an immediate more-value of the electronic platforms. As much for the suppliers as for the customers, this kind of platform, increases the option as much of purchase as of sale. The suppliers see the number of potential customers to increase - the prospection of new customers is facilitated. Customers can easily to compare prices of some suppliers and, because these are normally more competitive of what the ones that are not in the platform, can thus buy the lowest cost or in more advantageous conditions.

4 CONCLUSIONS

The technological advances of the last decades gave origin to an atmosphere where the organizations are forced to search new options to reduce costs, while at the same time to compete within their markets. This atmosphere needs flexible, capable and competitive organizations, able to make radical changes in the way they do business, employ people and use technologies (Varajão, 2001).

To be more competitive or, as we saw, to survive, companies, whatever is sector is - and the baking industry is not an exception -, they will have to anticipate, or at least to follow, the technological changes that happen everyday, in the national and international markets.

The Internet allows any company, big or small, to easily enter new markets, to conquer new clients, to establish relationships with new suppliers, and to establish new partnerships, without the material, geographical and time constraints of the conventional way to do business (Varajão, 2003).

The difficulty to find references about the baking industry in Portugal through an Internet search, for example using Google, suggests the need for the development of an Internet solution for this sector with vertical portal characteristics.

This kind of initiative, when it is well done, allows companies with a wider publicity and expansion of their markets, the keeping and attraction of new clients thought different and innovative ways, a better response to partners and clients, better services, new services available, and a cost reduction (in products, services and support).

An e-vortal for the baking sector can have a direct consequence on the growing of IT integration in business processes, making companies and organizations more in sync with current times and trends.

We intended with this work to give a contribution and to promote the development of this important industrial sector. The requirements presented in this paper should help the design and development of vertical portals for this sector.
REFERENCES


Vortalbuilding (2005b) What is a vortal?
ON EXPLORING CONSUMERS’ TECHNOLOGY FORESIGHT CAPABILITIES

An Analysis of 4,000 Mobile Service Ideas

Petteri Alahuhta, Pekka Abrahamsson
VTT Technical Research Centre of Finland, Kaitoväylä 1, Oulu, Finland
petteri.alahuhta@vtt.fi, pekka.abrahamsson@vtt.fi

Antti Nummiaho
VTT Technical Research Centre of Finland, Vuorimiehentie 3, Espoo, Finland
antti.nummiaho@vtt.fi

Keywords: Mobile service, technology foresight, user innovation, open innovation, lead-user innovation.

Abstract: Lead user driven innovation and open innovation paradigms seek to involve consumers and common people to innovative product development projects. In order to help developers choose ideas that meet the end users’ needs, we undertook a massive collaborative research effort and collected 40,000 ideas from 2,150 common people about future mobile services that they would like to use. We inspired each people to produce tens of mobile service ideas. In this paper we carry out an analysis for 4,000 ideas from the idea database. We had a particular interest in whether peoples’ ideas can be used in foreseeing the technology development needs. The results show that end users produce ideas that are conservative more than novel. Therefore, we claim that consumers’ technology foresight horizon is limited by the existing technological base. The second finding, linked to the previous one, is that the great majority of the ideas that consumers expressed could be realised utilizing existing technologies. The implication of this finding is that the idea database should be an interesting source of ideas for service developers. The third finding of the study, related to the methodology, is that a vast number of ideas can be collected fairly easily but analyzing them cost effectively is a challenge.

1 INTRODUCTION

End users of the products and services are instrumental for the success of any innovation. They will determine which products and services will ultimately become successful and which ones will fail. The end users’ market behaviour is difficult to forecast, and indeed history has shown that markets and end users may take unexpected turns. As an example, in the mobile service segment the high popularity of SMS (Short Message Service) led the operators to forecast that an MMS (Multimedia Messaging Service) offering would become an instant hit among mobile phone users as soon as the new technology was made available. For several reasons, the adoption rate of MMS has been significantly lower than the SMS service and in majority of the mobile service markets the SMS still keeps dominating the service offering despite of the opportunities that the newer technology offers.

Von Hippel (1986, 2005) has been one of the first authors to promote the use of consumers’ technological foresight capabilities to the fullest extent. In his seminal work on lead user driven innovation in 1986, he argues that in the area of high-tech product development lead users are actually ahead of the market place and possibly possess innovative ideas about the product or technology that they use. Von Hippel (2005) argues also that lead users may be quite willing to share their product related innovations freely and that these ideas are often commercially attractive.

Chesbrough (2003) popularized the concept of open innovation to distinguish from the traditional model of innovation, which he calls a closed innovation model. Chesbrough (2003) argues that external R&D (Research and Development) can create significant value and the internal R&D is needed to claim a part of this value. His open innovation model relies on building the business model first. He argues that companies should learn
to increase the effective usage of a company’s IP (Intellectual Property) either by selling it or offering it outside of the company’s boundaries by other means. Also, learning to acquire IP from companies, individuals and user communities (Baldwin, Hienerth & Von Hippel 2006) outside, whenever it fits the business model, bears significance in producing innovative products and services. In Chesbrough’s (2003) terms the open innovation model forms a new imperative for creating and profiting from technology.

Based on Von Hippel’s and Chesbrough’s concepts of lead user driven innovation and open innovation, the involvement of consumers in the development of future mobile services would seem to be of paramount importance to guarantee the continuous feed of innovative products and services. Little is known, however, to what extent the end users are capable of forecasting into the future (Un, Price 2007). Thus, the potential technological foresight capability of a mobile terminal user remains an area of research with little results.

When asked about future technology, we have a tendency to rely on forecasts made by information and technology research companies such as Gartner (www.gartner.com) and Standish Group (www.standishgroup.com/). Their predictions on future service development, builds upon technology. Markets and end-users are known to take unexpected turns. They may be even reluctant to adopt new technologies or services. Therefore it can be asked, how accurately are technology research companies able to predict future services, as the diffusion of services is dominated largely by other factors than those related to any particular technology.

In order to help companies and developers meet the needs of the growing number of mobile terminal users, we undertook a massive collaborative research effort and collected 40,000 ideas from 2,150 ordinary people about future mobile services that they would like to use. We inspired people not to think about the technology, but to express their concrete needs. Moreover, we did not request a single or few ideas from these people. Rather, we motivated people to come up with 30 or more ideas even if they would feel them to be either “silly” or not realistic in their minds.

In this paper, we report the results from a study where 4,000 ideas from the database were chosen for a technical scrutiny. We are particularly interested in understanding the technical foresight capability of an end user. We also wanted to know if by analyzing users’ ideas, one could identify specific technology development needs. Therefore, we paid particular attention to identifying possible technology related obstacles in front of successful mobile services, such as limitations in human technology interaction or in communication capabilities. As context awareness has been advertised as a technology to bridge the shortcomings of the mobile HCI (Cheverst et al. 2000, Korpipaa et al. 2006), we also analyzed to what extent the ideas suggest that this would be a desired development trend.

The paper is organized as follows: the next section outlines the research design of the study focusing on research methods, data collection and analysis. Also, the outline of the Idea Movement initiative that the present researchers took alongside with the principal high level results of the Idea Movement itself is presented. This is followed by the results of the analysis and a brief discussion of the implications for the mobile service business and research. The paper is concluded with final remarks.

2 RESEARCH DESIGN

This section discusses the issues related to research design. We will first introduce the research method and settings, and then we will briefly describe the data collection and means of analysis of the material. Finally, we will give some examples of collected ideas stored in the database.

2.1 Research Method and Settings

This study is an explorative research on the mobile service needs. The phases of an innovation process have been discussed in the literature (e.g., Koen et al. 2002). There is a strong need to better understand the role and potential of end users in foreseeing the technology development needs. The problem has also been discussed by Un & Price (2007). In order to explore the possibilities of ideas, we launched a national research project called the Idea Movement. The project started in the beginning of 2006. We set the goal to collect at least 35,000 ideas for mobile applications straight from the citizens, and to make these ideas accessible for everyone by publishing them in the Internet. Giving companies, organizations and individuals the opportunity to build on the ideas of thousands of people, we aim to accelerate the development and commercialization of new mobile services.

2.2 Data Collection and Analysis

Ideas were collected systematically in 31 workshops and events organized around Finland. Altogether
2,150 people participated in the events. The majority of the participants were university students, schoolchildren and elderly people. Workshops were also organized in schools, at workplaces and even in a shopping centre. Furthermore, there was a possibility to post ideas on the Idea Movement website, and about 200 people communicated their ideas in this way. Table 1 summarizes the division of ideas according to different age groups.

Table 1: Distribution of ideas in the Idea database.

<table>
<thead>
<tr>
<th>Age group (years)</th>
<th>number of ideas</th>
<th>Proportion of ideas</th>
</tr>
</thead>
<tbody>
<tr>
<td>7-12</td>
<td>1,800</td>
<td>4%</td>
</tr>
<tr>
<td>13-16</td>
<td>6,500</td>
<td>16%</td>
</tr>
<tr>
<td>17-19</td>
<td>4,300</td>
<td>10%</td>
</tr>
<tr>
<td>20-50</td>
<td>21,000</td>
<td>51%</td>
</tr>
<tr>
<td>50-90</td>
<td>4,300</td>
<td>10%</td>
</tr>
<tr>
<td>Unclassified</td>
<td>3,100</td>
<td>8%</td>
</tr>
<tr>
<td>Total</td>
<td>41,000</td>
<td>100%</td>
</tr>
</tbody>
</table>

Each workshop was kicked off with a short introduction to idea generation techniques, followed by brainstorming sessions both individually and in groups. At the beginning of each event we asked the participants to produce ideas about a given topic, but we also accepted ideas that did not match the original subject of the particular brainstorming session. We instructed each participant to produce 20 ideas individually and then form groups of 3-4 people. The ultimate goal of idea generation was that each group would deliver together 100 ideas, or more.

Ideas are 1-2 sentence descriptions of a mobile service idea or an expression of a need that they think could be fulfilled utilizing mobile technology. The progress of a brainstorming session has been described in more detail by Leikas (2007).

Idea analysis was carried out in two phases. First a group of five reviewers did a qualitative analysis of 2,000 ideas so that we got an understanding of the potential of the ideas. According to the reviewers’ subjective opinion, each idea was classified into one of the following four categories: Excellent, Interesting, Conventional and Not a mobile service idea.

We also carried out an analysis of the technical aspects of the collected ideas in order to get an understanding of the technologies that are required for implementing these ideas. Also, the essential constraints that may currently prevent the implementation were considered.

For the analysis three technology areas (access technology, context information, human-technology interaction) and major technical constraints were identified. Furthermore, we explored what kinds of integrated or interconnected devices were proposed to be part of mobile services. The analysis was carried out by reviewing a selected portion of the ideas and categorizing them based on the defined categories. Each idea was classified based on its most obvious way of implementation. For this paper, we selected 4,000 ideas (i.e., 10% of all ideas) for a detailed inspection. The selected dataset consists of 2,000 ideas produced by high-school students (the age group of 17-18 years) and 2,000 ideas from university students and employees (the age group of 20-50 years).

2.3 Data Examples from the Idea Base

The themes of covered a large variety of different topics including Public transportation, Travelling, Work and Learning, Culture and entertainment, Hobbies, Wellbeing and health, Shopping and service, Family, Friends and relatives, Household management and living and Everyday activities.

In order for the reader to comprehend the nature and type of the mobile service ideas, we have listed below some ideas classified to the four suggested categories: Excellent ideas, Interesting ideas, Conventional ideas and Not a mobile service idea. These categories were created in an early phase of the analysis in order to quickly have an exploratory view of the ideas in the database.

Examples of ideas classified as Excellent are:
- Warning if parking time is running out. Possibility to automatically buy or order additional parking time.
- Location-based filtering of incoming calls. E.g., No work-related calls at home.
- Locating of friends (if they allow).

Examples of Interesting ideas are:
- In a rally, one can get information about the condition of a car. A spectator will know if the car is going to break down.
- Service that tells you how strong the punch in your glass is.
- Opportunity to order "good explanations" when coming home late.

Examples of Conventional ideas are:
- The controlling of a home automation system by a mobile handset. Being able to switch on/off lights, heating, etc.
- Message from library if a new book by your favourite author is available.
- Service for getting current hit music to your mobile phone.
Finally, examples of ideas that are *not a mobile service idea* include:

- Mobile phone making a cup of coffee
- 3” nails inside the mobile phone
- Rat trap

### 3 RESULTS

In this section we will describe the results of the technical analysis of 4,000 ideas. At first the application categories are introduced, followed by technical constraints in five major areas suggested by the literature.

#### 3.1 Application Categories

We first classified mobile service ideas into eight distinct categories based on the idea clusters identified in the data and had one category for miscellaneous ideas. The results have been summarized in Table 2.

<table>
<thead>
<tr>
<th>Category</th>
<th>Description</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Information pull</td>
<td>Retrieving information for some purpose (possibly based on location).</td>
<td>30%</td>
</tr>
<tr>
<td>Information push</td>
<td>Receiving information automatically (possibly based on location).</td>
<td>14%</td>
</tr>
<tr>
<td>Locating (persons / objects)</td>
<td>Locating or following some (nearest) person or object.</td>
<td>9.2%</td>
</tr>
<tr>
<td>Communication</td>
<td>Social discussion channel.</td>
<td>7.0%</td>
</tr>
<tr>
<td>Service request</td>
<td>Ordering a personal service (possibly based on location).</td>
<td>4.7%</td>
</tr>
<tr>
<td>Content production</td>
<td>Producing content.</td>
<td>4.3%</td>
</tr>
<tr>
<td>Payment</td>
<td>Using mobile device as a means of payment.</td>
<td>4.0%</td>
</tr>
<tr>
<td>Identification</td>
<td>Using mobile device as an identification device.</td>
<td>3.4%</td>
</tr>
<tr>
<td>Others</td>
<td>Applications that do not fit into other categories.</td>
<td>24%</td>
</tr>
</tbody>
</table>

According to technical analysis people see mobile services largely as an information channel, which can be used when ever they need to know something. 30% of ideas fall into this category. Significantly fewer ideas proposed automatic information delivery to users’ terminals. Only 14% of the ideas proposed push services such as advertisement. Locating missing objects and persons represented 9.2% and communicating, for example, with social communities 7% of ideas. Service requests (4.7%), mobile content production (4.3%), payment (4%) and identification (3.4%) represented a perhaps surprisingly low popularity in the analyzed set of ideas. 24% of the ideas fell into the category others. We think this indicates that there is a versatile set of activities in everyday life where mobile technology could provide value to the users.

#### 3.2 Technical Constraints

In this paper we want to explore the nature of user innovation ideas from the technical viewpoint. We are particularly interested in what kinds of technologies are needed to realise the ideas that users have proposed. More specifically we are interested in:

- a) what seems to be the main constraints for services not being developed,
- b) what communication and access technologies are required for the ideas,
- c) what kind of context information, if any, is required for services that users propose,
- d) what are the challenges in Human Technology Interaction (HTI) technologies, and
- e) what kinds of integrated or interconnected devices are required for the ideas.

##### 3.2.1 Analysis of Major Constraints for Adoption of Mobile Services

One of the main findings of our technical analysis is that there are no major technical constraints in developing most of the mobile service ideas into actual mobile services for people. However, we can pinpoint some constraints that are quite typical for mobile services and well known by the literature.

Even if these constraints are not immediate roadblocks for services, they weaken the user experience so much that eventually users may not start using the services in the first place or they stop using the services because of the poor experience.

In Table 3 technical factors that may hinder user experience have been summarized. Small screen size and low bandwidth are perhaps the most important constraints in mobile services. Other factors that make it difficult to realise some of the ideas are low processing power, high power consumption of the terminals and a limited amount of memory available in the terminal.

Some of these constraints will be overcome in time. Such, such as bandwidth, processing power and amount of memory. Screen size and power...
consumption, however, are harder obstacles for developers. The trend towards extensive multimedia communication requires as large a screen as possible and higher processing power that leads to higher power consumption.

Table 3: Major technical constraints identified in collected mobile service ideas.

<table>
<thead>
<tr>
<th>Category</th>
<th>Description</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Screen size</td>
<td>The user interface requires a large screen.</td>
<td>3.5%</td>
</tr>
<tr>
<td>Bandwidth</td>
<td>A lot of network traffic occurs.</td>
<td>2.9%</td>
</tr>
<tr>
<td>Processing power</td>
<td>Processing power is especially important.</td>
<td>1.0%</td>
</tr>
<tr>
<td>Amount of memory</td>
<td>A lot of multimedia or other memory expensive data is handled.</td>
<td>0.96%</td>
</tr>
<tr>
<td>Battery duration</td>
<td>A lot of power is needed.</td>
<td>0.91%</td>
</tr>
<tr>
<td>Touch screen</td>
<td>The user should be enabled to interact with the application by touching pictures or words on the screen.</td>
<td>0.32%</td>
</tr>
<tr>
<td>Keyboard</td>
<td>The user must type lots of text.</td>
<td>0.30%</td>
</tr>
</tbody>
</table>

It is also to note that the total percentage of ideas having major technical constraints is very low. In other words, technology is not the main constraint in implementing most of the ideas.

3.2.2 Analysis of Required Communication and Access Technologies

In the technical analysis we wanted to consider what communication and/or service access technologies are required for realizing the ideas. Table 4 summarizes the analysis of required access technologies.

In the table we can see that the vast majority (67%) of foreseen services can be realized utilizing existing access technologies such as GSM/GPRS (2G) or 3rd generation cellular networks (3G). 13% of ideas required short-range communication solutions, such as Bluetooth or Near Field Communication (NFC). Quite few services really required such communication technologies as satellite communication, cell-casting or a combination of different communication technologies.

During the time of collecting the ideas (in 2006) there were quite lively debates about the promised break-through of mobile-TV. Our study did not support the need of these technologies and it seems that users have not adopted to broadcasting-type access technologies beyond FM-radio-receivers in their mobile handsets.

Table 4: Required access technologies.

<table>
<thead>
<tr>
<th>Category</th>
<th>Description</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>3G</td>
<td>Text and multimedia based services.</td>
<td>67%</td>
</tr>
<tr>
<td>2G</td>
<td>Mainly text based services.</td>
<td>60%</td>
</tr>
<tr>
<td>Short-range communication</td>
<td>Communication over a few meters or by touch. (Bluetooth, NFC)</td>
<td>13%</td>
</tr>
<tr>
<td>WLAN</td>
<td>Mainly indoors or in areas with lot of people.</td>
<td>1.5%</td>
</tr>
<tr>
<td>Broad-casting</td>
<td>Sending same information to all cell phones.</td>
<td>1.1%</td>
</tr>
<tr>
<td>Satellite</td>
<td>Worldwide access may be needed everywhere including mountains and seas.</td>
<td>0.47%</td>
</tr>
<tr>
<td>Combination of two or more access technologies</td>
<td>Non-trivial combination of multiple access technologies.</td>
<td>0.44%</td>
</tr>
<tr>
<td>Cell-casting</td>
<td>Sending same information to all cell phones within a base station.</td>
<td>0.22%</td>
</tr>
</tbody>
</table>

3.2.3 Requirements for Human Technology Interaction (HTI) Technologies

Earlier in this paper we discussed the major technical constraints when considering the implementation of mobile service ideas.

Table 5: Human Technology Interaction technologies required in mobile service ideas.

<table>
<thead>
<tr>
<th>Category</th>
<th>Description</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Image / Video analysis</td>
<td>Extracting information from image or video.</td>
<td>1.8%</td>
</tr>
<tr>
<td>Speech synthesis</td>
<td>Ability to produce sound that resembles human speech.</td>
<td>0.86%</td>
</tr>
<tr>
<td>Augmented reality</td>
<td>The idea that an observer's experience of an environment can be augmented with computer generated information.</td>
<td>0.76%</td>
</tr>
<tr>
<td>Speech recognition</td>
<td>The ability to recognize and carry out voice commands or take dictation.</td>
<td>0.69%</td>
</tr>
<tr>
<td>Audio analysis</td>
<td>Extracting information from audio.</td>
<td>0.49%</td>
</tr>
<tr>
<td>Gesture recognition</td>
<td>The ability to recognize human gestures, usually hand motion.</td>
<td>0.22%</td>
</tr>
<tr>
<td>Gait pattern recognition</td>
<td>The ability to recognize gait.</td>
<td>0.02%</td>
</tr>
</tbody>
</table>
HTI-technologies do not seem to be a major obstacle in the track of creating mobile services. However, we know that poor usability and restrictions in HTI-technologies weaken the user experience (Hartmann, Angeli & Sutcliffe 2008). Developers are investing lots of effort creating solutions that make mobile interaction more intuitive. We wanted to look at a number of HTI-technologies in order to see if end users generating service ideas are in need of novel HTI technologies. Table 5 summarizes the analysis.

Based on the analysis we can say that users either can not imagine or do not seem to hunger for services with novel HTI capabilities, such as speech synthesis, augmented reality, speech technologies or gesture recognition.

3.2.4 Analysis of Context Information Required for Services Ideas Users Proposed

The use of context information in mobile services has been an active research discipline for already a decade. Researchers and developers have proposed a number of different context-aware mobile services and applications.

Table 6: Context Information required in mobile service ideas.

<table>
<thead>
<tr>
<th>Category</th>
<th>Description</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Location</td>
<td>Location of the user or some object.</td>
<td>25%</td>
</tr>
<tr>
<td>Time</td>
<td>Time that is relevant in a non-trivial way.</td>
<td>4.7%</td>
</tr>
<tr>
<td>Activity</td>
<td>What the user is doing.</td>
<td>1.8%</td>
</tr>
<tr>
<td>Identities of nearby people</td>
<td>Identities of people that are close-by.</td>
<td>1.8%</td>
</tr>
<tr>
<td>Emotion / Mood</td>
<td>How the user or some other party feels.</td>
<td>1.6%</td>
</tr>
</tbody>
</table>

Our analysis, summarized in Table 6, confirms the fact that location seems to be the most valuable and versatile context information to be used in mobile services. Even 25% of service ideas require location information. Time (4.7%) is another obvious context information. Identifying the activity of the user, the social context (people nearby) or an emotional situation do not seem to be as important factors for the proposed mobile service ideas.

3.2.5 Analysis of Needs for Integrated or Interconnected Devices

We also wanted to look at what kind of integrated or interconnected devices people wanted to have in their mobile terminals. The results are summarized in Table 7.

Table 7: Needs for integrated or interconnected devices.

<table>
<thead>
<tr>
<th>Category</th>
<th>Description</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Meters/Sensors</td>
<td>Measuring a property.</td>
<td>6.8%</td>
</tr>
<tr>
<td>Controllers</td>
<td>Controlling a device.</td>
<td>3.3%</td>
</tr>
<tr>
<td>Others</td>
<td>Mobile phone being or using other devices.</td>
<td>9.9%</td>
</tr>
</tbody>
</table>

In the analysis, we identified two clear categories, namely Meters/Sensors and Controllers. The first category stands for ideas that have some kind of sensing or measuring functionality. Ideas in this category include, for example, a mobile phone with a step counter or an alcometer. The second category includes ideas related to controlling other devices using a mobile phone, for example, the possibility to open a home door for someone from distance using a mobile phone. In addition to these categories, we found a versatile set of ideas proposing that a mobile phone includes or uses another device, such as a laser pointer or a projector.

4 DISCUSSION

Literature claims that people may have unique, interesting and potentially commercially attractive ideas about technologies (Von Hippel 2005). Literature also holds that bright ideas may emerge from inside an organization as well as from the outside (Chesbrough 2003). We challenged these arguments and collected a vast amount of mobile service ideas from students, working-age people, school-children and elderly people. We argue that an in-depth analysis of this material will provide us valuable insight about users’ everyday needs and wishes concerning mobile and ubiquitous technologies and services.

In this paper we were particularly interested if user-generated mobile service ideas could be used as a tool in foreseeing the need for technological development. The reasoning behind this idea is as follows: if users would propose needs or ideas that cannot be realised using existing technologies, it would create a potentially attractive target for technology developers.

The discussion is organized around three practical implications of the study. First, we claim that it is relatively simple to collect a large amount of data about people’s ideas or needs. The real challenge is to analyze these ideas cost effectively. Second, we claim that the adoption of a vast variety of mobile services is not primarily limited by missing technologies. In fact, the majority of all ideas can be realised with existing technologies. And third, we claim that people are not likely to propose
very futuristic service ideas. On the contrary, they are tightly limited to existing understanding of the capabilities of mobile systems. We could say that people are hindered by the technological frame of today.

4.1 Consumers Freely Share their Ideas

During the course of the Idea Movement project we have seen that collecting a large amount of idea material is not an overwhelming task. People were willing to participate and reveal their ideas in the brainstorming sessions that we organized in various schools, universities and companies. In these events we offered participants refreshments, such as coffee and snacks, but the participants were not offered any financial or material rewards for participation. This seems to confirm the claims that economic factors are not the primary source of motivation for people who share their ideas (Lüthje 2004). In the beginning of the process of collecting these ideas we thought that IPR and ownership of the ideas would become an important topic of discussion. Surprisingly, only a few people raised the issue. Perhaps this was due to the policy to publish all the ideas, and the fact that these ideas are mainly very short descriptions of a need or service – not detailed business ideas.

Collecting a large amount of ideas is not as difficult as it may sound. Instead (cost) effective processing of these ideas is much greater challenge. We have tried various automated systems for analyzing the data, but they do not seem to work very well. This is mainly due to the free format of ideas. The same idea can be expressed in various ways with different kinds of language, i.e., using standard language, dialect, spoken language or even slang.

An interesting avenue to pursue for processing these ideas is to distribute the workload to large groups of people using the Internet. In this approach persuading and motivating people to work on the ideas requires further consideration. We have done some experiments on carrying out Internet based analysis, but the results fall out of the scope of this paper and thus will be published in the future.

4.2 95% of the Ideas can be Implemented

In our study it became evident that the vast majority of mobile service ideas can be implemented using existing technologies. Technical components of mobile services are typically a mobile terminal (mobile phone) with a browser or a dedicated application, wireless network access and a server with service specific functionality.

According to our analysis only 5% of proposed mobile service ideas had major technical obstacles. One such technical obstacle is, for example, small screen size, which is particularly problematic in navigation and multimedia services. Another technical challenge is limited bandwidth of mobile access. Bandwidth limitation causes troubles in multimedia intensive services such as video-conferencing and multimedia streaming. Power consumption and battery technology may also limit possibilities in few service ideas.

The development of technologies takes some of the current obstacles into history. Some of the new mobile terminals from various vendors have much larger displays than a couple of years ago. Also, new network technologies are provided with improved bandwidth and the computing power will increase, as well as the size of memory.

In fact, we could argue that all proposed mobile service ideas could be realised using existing technology, but the user experience and cost structure might not be quite satisfactory for commercial deployment.

4.3 Consumers’ Technology Foresight is limited to Existing Technologies and Paradigms

While we suggested that approximately 95% of all ideas can be realised with existing technologies, it can be said that this method and project produced a large amount of valuable data for companies in quest of new services right now. Even though researchers and developers may be interested in technology forecasts, this material may not readily reveal the future developments of mobile technologies. Instead, we suggest that extensive pre-processing may enable the discovery of novel patterns not yet identified. We should try to identify some “weak signal” phenomena from the set of ideas instead of stronger trends. These stronger trends tend to be the conventional ideas that are familiar to everyone.

Conservative ideas may be due to the lack of consumers’ understanding on the new possibilities of mobile technology or that the majority of users are so tied with their current context that they cannot imagine new paradigms and revolutionary ideas. We are far from being disappointed with the evidently short technology horizon of people. When we consider these ideas from a business point of view, we see that the ideas deemed Conventional represent needs of a large group of potential customers that
have not been met by the service providers. We see potential breakthrough opportunities by new offerings formed from thousands of raw ideas.

We were also expecting to see more ideas related to new developments in the Internet such as social networking and the creation of mobile user-generated content. Another topic that we expected to see more was commerce, payment and identification of users using mobile technology. Both of these developments are active in the Internet, but participants of these events did not see the need for carrying out these tasks with their mobile systems.

Many of these services, however, require that new revenue-models should be defined, which would enable the penetration of mobile services based on the content and interest rather than merely on data traffic costs.

5 CONCLUSIONS AND FUTURE WORK

The 40,000 ideas for new mobile services expressed by the participants are far from well-defined business concepts ripe for commercialization. The ideas still require further development and professional elaboration before they can be introduced to the market as products or services.

In this study we presented the findings based on a technical analysis of 4,000 ideas. The findings were grouped into three principal practical implications: 1) Consumers share their ideas freely, 2) vast majority of the ideas can be implemented by means of existing technologies, and 3) consumers' technology foresight appears to be restricted by their experience with current technology and paradigms.

In order to advance this development we plan to carry out a set of analyses for the idea database. One particular topic that we are interested in is comparing the differences in ideas of people of different age. We want to explore this further in future studies. We will continue to expand the Idea Movement to other countries, cultures and nationalities in the near future. The Idea Movement’s (www.idealiike.fi) idea bank is now open in Finnish for all commercial, research and educational purposes. The ideas will be opened up in English as well. This enables the development of these ideas in global context.

We predict that the actual value of these ideas is generated in the idea refinement phase where hundreds of ideas are combined and enriched through the conceptualization process. This is a task reserved for companies and organizations with the ability to efficiently commercialize the services. Idea Movement, therefore, makes a link between the intellectual capital of ordinary citizens and the technology business know-how of companies which results in a potential win-win situation.

ACKNOWLEDGEMENTS

We acknowledge Idea Movement Partners, Maaretta Törrö and Anssi Öörni.

REFERENCES


AN ANALYSIS OF CONTEXT-AWARENESS 
IN COMMERCIAL MOBILE SERVICES

Ana M. Bernardos, Daniel Marcos and José R. Casar
ETSI Telecomunicación, Universidad Politécnica de Madrid, Ciudad Universitaria s/n, 28040, Madrid, Spain
{abernardos, dmarcos, jramon}@grpss.ssr.upm.es

Keywords: Mobile data services, context-aware applications, location based services, mobile business.

Abstract: This contribution aims at analyzing context-awareness from a commercial point of view, studying how the utilization of context descriptors and features in mobile services has evolved during the period 2003-2008. The analysis is based on the information collected during a Technology Watch activity; this methodology has provided us with a large database of mobile services, built from monthly updates of mobile novelties and commercial launches. Services in the database have firstly been categorized regarding their functionalities, in order to get the big picture of the mobile ecosystem in this period of time. Afterwards, we have identified a list of descriptors (personal, physical or activity related) and features (related to resources discovery, management and communications and also to advances HCIs) which are usually identified in context-aware applications and systems prototypes. The use of these descriptors and features has been evaluated for each service and some trends have been detected. Our general conclusion is that nowadays few commercial mobile services can be considered “context-aware”, although isolated features (mainly related to personalization) are perceived in many applications. Due to their functionalities, location aware services and mobile social networks are leading the use of context parameters.

1 INTRODUCTION

Simultaneously with the evolution of commercial mobile technologies and markets, a lot of research has focused on bringing to reality the paradigms that are supposed to sketch the future of mobility (Mohr, 2008). Ubiquitous and pervasive computing (Weiser et al., 1999) inspire the concept of ambient intelligence (ISTAG, 2001), a smart and sensitive scenario where context-aware applications (personalized and status adaptive services) make easier everyday activities (Schilit and Theimer, 1994). 

To date, context-awareness has inspired many research prototypes in different application fields (museums, airports, hospitals, smart homes, etc., see Section 2). From an analysis on the evolution of commercial mobile services during 2003-2008, this contribution aims at shedding some light on how context-aware research concepts and features are being transferred into market applications. Our final motivation is to identify the hindrances (technical, privacy and business related issues) which need to be overcome to generalize the use of context information in commercial mobile services.

It is important to remark that this is neither a forecasting work nor a study on adoption and pattern of use study on mobile services. Both types of analysis are already common in (academic and corporate) literature. With respect to forecasting, there are some broad-scope proposals that aim at predicting the penetration and the traffic in cellular mobile networks –eg. (Arvidsson et al., 2007)- and also studies focused on niche services or particular markets –eg. (Funk, 2007). On the other hand, mobile services adoption is attracting great interest: for example, Kelleher (2007) analyzes four studies on this issue, Bouwman et al. (2006) goes deeper on barriers and drives that condition the services’ use and Bina and Giaglis (2005) and Gilbert and Han (2005) analyze users’ preferences taking into account their life styles, needs and demographical characteristics. Verkasalo (2007) studies the users’ preferences from the analysis of the data traffic generation of a group of representative applications.

Using a complementary perspective to mobile data adoption studies, our contribution is firstly conceived to analyze the generation of mobile services. With this purpose, during the last five years (2003-2008), we have followed a Technology Watch methodology which has provided us with a monthly updated data base of new mobile services. In this contribution, we elaborate on the long series of
gathered data in order to show how mobile services and context-awareness presence has evolved.

The paper is organized as follows. Section 2 goes deeply into the concept of ‘context-awareness’ and gathers a review of its areas of application. Section 3 presents the methodological approach used in this study. From the empirical analysis, Section 4 provides a view of the mobile ecosystem, to frame Section 5, which elaborates on the level of “context-awareness” that current mobile applications have. Section 6 concludes the paper with some open issues for the generalization of context-awareness in commercial mobile services.

2 FROM LOCATION BASED SERVICES TO CONTEXT-AWARE APPLICATIONS

From the late nineties on, location has been the enabler of a number of commercial “precontext-aware” services, such as family finders, location based advertising, area billing, pervasive games, trackers, real time location systems (RTLS), etc. (Bernardos et al., 2007). After a complicated take off of LBS (Kaasinen, 2002), standalone navigation and tracking applications have nowadays become popular, partly due to some device manufacturers’ efforts to promote the use of GPS enabled devices. Meanwhile, innovative mobile services related to Web 2.0 have shown up (it is the case of mobile social software or applications making easier content geotagging, for example) and some analysts expect them to contribute to widen the adoption of LBS.

Simultaneously to LBS evolution, research in context-aware systems and services has gone ahead. Since the pioneers Active Badge or PARCTab projects, advances and challenges in positioning techniques, semantic context representations or software architectures for context-awareness have evolved and a great variety of application environments have been explored. By way of illustration, following there is a short list of some of them:

- **Context-aware mobile guides**: many initiatives (such as the breaking Cyberguide or GUIDE) have focused on the development of tourist context-aware mobile guides. Museums (e.g. Sotto Voce or Exploratorium projects) and exhibition centres (e.g. Hippie prototype, mExpress project or XGuide application) have also been inspiring environments. Augmented reality techniques are nowadays being combined with context-awareness to achieve a new user experience.

- **Productivity applications for working spaces** such as classes, campuses (e.g. ActiveCampus, Classroom 2000 or eClass projects) and offices (e.g. Context-Aware Office Assistant).

- **Fieldwork applications** for environments where recording and filtering data are prior tasks, such as laboratories (e.g. Labscape) or archaeology areas (e.g. the Context-aware Archaeological Assistant).

- **Smart homes** are highly pervasive scenarios with sensors, actuators, wireless networks, etc. (Meyer and Rakotonrainy, 2003); in particular, there is a growing interest in applications aimed at supporting daily living activities and well-being (especially of elderly and disabled people) (e.g. Wireless Wellness Monitoring and Howel projects).

- **Health care environments**, such as hospitals - Bricon-Souf and Newman (2007) contains a survey- or even operating rooms (Agarwal et al., 2007).

- Horizontal applications such as memory aids, location annotation software (e.g. GeoNotes) or context-aware telephony (callers are provided with context information about the receivers (Khalil and Connelly, 2006)).

- Other challenging areas of application are context-aware mobile learning (e.g. Mobilearn project), context-aware mission critical support (e.g. Siren project for firefighting) or ubiquitous mobile gaming (e.g. Botfighters).

Context-aware services share requirements with general mobile ones (eg. with respect to usability and interaction mechanisms, device requirements - screen sizes or power consumption-, network speed, etc) but, the same as location based services, are especially dependant on:

1) **Stable location mechanisms.** Location is a relevant descriptor of context that is used as a filter in most context-aware applications. The quality of the location estimation has a direct impact on the user experience. So, reliable, transparent and latency-controlled location mechanisms are still needed to offer an acceptable user experience. To date, GPS (and its variants, such as A-GPS) is the most used positioning mechanism outdoors. But GPS does not perform well indoors, where the multiplicity of networks make possible to configure diverse solutions. Roaming among different positioning technologies is a challenge to be solved.

2) **Reliable privacy management.** Context-awareness means acquiring and handling identity, real time location or activity data. Users must be
aware of which personal data are being managed and have control over the acquisition mechanisms. The combination of security, pseudonym use or cloaking zones may confer the user with a sufficient feeling of control.

3) Sustainable business models: clarifying the “ownership” of context data. Mobile operators are nowadays handling a lot of personal information (such as rough cell-based location data) and providing third partners with the tools to access it. At the same time, hardware manufacturers are equipping mobile device with autonomous sensors (eg. GPS or NFC readers), capable of acquiring context information directly from the device. Some of them are also partially shifting their revenue sources to the provision of applications that use the acquired parameters. These two different approaches to context-data acquisition may determine the development of context-aware services.

“Fully-compliant” context-aware applications seem to be in a prototype stage yet. But some mobile services are already intensively using some context descriptors. In the next pages, we intend to analyze to which extent context-awareness is implemented in current commercial mobile services and how it has evolved in the last few years.

3 METHODOLOGICAL APPROACH

Technology Watch, according to the Standard UNE 166006-EX, is the continuous process of obtaining and selecting information from a scientific and technological environment, in order to turn it into useful knowledge by analyzing and spreading it, so that organizations can take advantage of it and make better decisions by being ahead of changes.

From 2003 to date, a Centre for Technology Diffusion in the UPM, CEDITEC, accomplishes a Technology Watch Programme in ‘Wireless Technologies and Mobile Applications’, which includes the generation of a monthly newsletter and a technology watch report on this issue. The watch process starts with the selection of critical information sources related to mobility, more than 60 sources ranging from technical publications to mobile market focused weblogs, also including general press. Afterwards, a methodical process of data collection allows expert filtering to elaborate a selection of the most relevant information in the month: news, reports, publications, patents, products and also an average selection of the 15 newest and most representative mobile services found in that period of time.

This way, our time-indexed database gathers information of 735 mobile services, from May 2003 to March 2008. As the technology watch process has not been to collect every new service, but the most representative or new, the quantitative analysis happens to be done on critically filtered data. Besides, the services are mainly focused on European and North America markets (as it is difficult to directly access Asian information sources, due to language restrictions).

In this contribution, the services’ information has been submitted to further analysis, considering the aspects of user experience, technology evolution, business case, and contextual functionalities developed in the following sections.

4 BUILDING THE BIG PICTURE OF THE EVOLUTION OF MOBILE SERVICES

In this section, we comment on the mobile ecosystem evolution between 2003 and 2008, the considered period of analysis. The objective is to provide the reader with a general (not complete) picture on how the situation has changed in the last five years, supporting our statements with some results of the analysis of the collected data.

a) Entertainment and self content generation together with information services has driven the production of mobile applications.

To analyze what the present offer of mobile services is and how it has evolved during the last five years, we have inferred a functional classification on the stored data. It is composed of 12 categories; we consider that a potential category becomes a formal one when it groups 20 services as a minimum. The classification is as follows (note that some services may match more than one category):

- Location and tracking: of people or mobile assets (94 services).
- Information: news, weather reports, etc. in addition to eLearning and language translation services (130).
- m-commerce: mobile payments, financial services and m-ticketing (62).
- Browsing and searching: in the Internet (41).
- Connectivity and communications: including VoIP, SMS, MMS, IM or videoconference (96).
- Applications and productivity: e-mail applications, document formatting, and remote access to computers or mobile phones (47).
- Security and safety: data security and surveillance applications (20).
Entertainment and self content generation: photos, video, music, and games downloading, creating, publishing and sharing (155).

Personal information management: calendars, reminders, customization, settings (78).

Social networking (49).

Mobile experience enhancers: new interfaces, content adaptation, voice to text conversion, 2D codes reading (72).

Miscellaneous (32).

Most categories include some subtypes of services which share the final objective but differ in their functional implementation. It is the case of multimedia content and information services. Both have driven the commercial production of mobile applications in the last years (accounting for 21% and 18% over the total).

The other main categories are connectivity and communications (13%), location and tracking services (12.7%) and personal information managers (10.6%).

b) Towards mobile 2.0: downloading, but also sharing and uploading contents.

“Mobile 2.0” is the way to refer to the convergence between the social web (or Web 2.0) and the basics of mobility (personal, localized and always-on). From a practical point of view, Mobile 2.0 is about connecting your phone to download your favourite podcasts, read your RSS feeds, do a one-click image upload to an on-line photo management and sharing application, consult the location map while on the road, tag your streamed videos or update your moblog.

It is noticeable that during 2003-2005, the content offer was mainly focused on data downloads, whereas from 2006 on, an increasing trend in services supporting the creation and sharing of mobile content is detected. At the same time, the presence of other categories, almost undeveloped before, are intensified: it is the case of browsing and searching or doing social networking.

c) A technological view: enhanced networks and sensing mobile devices.

There are two key points regarding technology that shape the changes of the mobile ecosystem during 2003-2008: a) mobile devices have evolved from traditional mobile phones to smart devices with increasingly embedded technology and b) the communication infrastructure is not based on cellular technology any more. Cellular communication networks have enhanced their performance and availability, but alternative technologies have shown up and increased their penetration rate.

With respect to mobile devices features, Figure 2 illustrates an approximated timeline on how new technologies have been embedded into mobile devices. The arrows point out the early introduction of a technology in a commercial phone in the general market, although its technical availability may be previously detected and its availability as a commercial common feature may not have happened until several months later.

On the other hand, a very general picture of communication infrastructures shows that:

- Cellular networks have increased their data rate, through UMTS/WCDMA technologies (384kbps) and its enhanced versions: HSDPA (14.4Mbps), HSUPA (5.6Mbps), HSPA+ (42 / 11.5Mbps) and the next LTE – 3GPP Release 8 (>100 / >50Mbps).
- Growing production of standards: WiFi (IEEE 802.11g, June 2003; next IEEE 802.11n in 2009); WiBRO (TTA, late 2005), WiMAX (ITU, October 2007), LTE (GSA, January 2008).

Figure 1: Evolution of content services (percentage of services within the total of each quarter).

Figure 2: Approximate timeline of technologies in mobile devices.
Great take off of WiFi networks, with 89% growth of WiFi business Hotspots (those located in airports, hotels, retailers, etc.) over H2 2007 (iPass, 2008).

Fixed to Mobile Convergence (FMC) has been a well-known strategy of some operators, adopting UMA and IMS in order to provide better indoors mobile coverage.

Enhancements in location technologies: E-911 in USA, market-driven deployment of location based services in Asia-Pacific zone (DTI, 2004), growing integration of GPS receivers in mobile devices all around the world.

d) Business: mobile service providers shift from operators towards software companies.

Generally, the development and commercialization of a mobile service implies the interaction of a number of stakeholders. Obviously, the value chain (or network) composition is very dependent of the final service. For example, location based services value chains are generally complex: in the case of an A-GPS mobile location server, it is possible to identify at least ten roles doing their part to build the final service (Bernardos et al., 2007), and it will be the mobile operator the one offering and billing the service. On the other hand, a Bluetooth based social application, offering discovery of peers and communications will only require an application provider and, indirectly, a handset compatible provider. In this case, the user will get the application directly from the service provider website, and will be not charged for using it.

Figure 3: Comparison of mobile services’ providers.

On the whole, operators start to change their walled-garden strategies (Wieland, 2007) towards more open ones, in part to incentive data traffic through popular 2.0 applications and to let the users browse in a more familiar way. Well-known web services providers that have mobilized their applications (p.e. Google, Yahoo! or Skype) are acting as users’ attractors to mobile data services.

From our analysis of data, we get the following picture (Figure 3): mobile operators started dominating the provision of content and services to mobile devices, whereas in the last six quarters this market has surrendered to Internet software service providers.

5 CONTEXT-AWARENESS IN COMMERCIAL MOBILE SERVICES

5.1 A List of Context-Aware Features

From the classification of mobile services proposed in Section 4.1, the reader might conclude that many of the considered services are not context-aware at all. But some of them—although far from being truly context-aware applications—could have some ‘context-aware’ features; intuitively, it is the case, for example, of those services grouped into the ‘location and navigation’ or ‘social networks’ categories.

Measuring the context-awareness level of a mobile service is not an easy task. For our analysis, we have identified a number of basic questions that a context-aware service should be able to answer:

1) **Who the target entity/user is and what does it/he/she want/like?** Some services will need to handle personal descriptors, these understood as identity, profile, preferences and group membership related issues. Personalization and services’ adaptation to the user’s profile are key for many applications.

2) **Where the user is?** Location is a physical descriptor that usually provides significant knowledge about the user’s context. Location acts as a situational filter that is influenced by the positioning system accuracy. Being the basis of navigators and trackers, location availability is making possible some innovative services such as mobile locative social software, pervasive games or geotagged content making (photos, blog posts, etc. may be georeferred).

3) **How the user is?** Environmental and biometric data use. Apart from location, context-aware services may use other physical descriptors to describe the environment where entities are plunged (this may be useful, for example, for environmental and agriculture surveillance applications and domotics), or even to monitor their biometric parameters, in order to infer information about the target entity physical state.

4) **What is the current user’s activity?** Not independent of the previous issue, information about the current activity and logical state (presence or connectivity, for example) of a target entity may
complete the inferences about its situational condition. Features such as presence announcement and state dissemination or availability notifications are considered in this point.

5) What does the user intend? This feature is related to personal or professional activity data gathering, under the shape of calendars, schedules or notifications, or behavioural patterns.

6) Who/what is near the user? In certain applications, building a social group based on proximity issues with defined privacy levels and collaboration policies is the service’s leitmotiv. Context-aware services combine physical and situational data with the management of preferences and interests, creating social networks which promote direct interaction among peers. Apart from discovering people, discovering and interacting with daily objects through wireless location (WiFi or BT based systems, for example) and proximity technologies (such as NFC or RFID) is also possible.

7) With whom the user may communicate? Applications which are focused on enabling peer communications make possible the formation of proximity-based groups.

8) How are the user’s interaction mechanisms? Context-awareness aims at facilitating daily living and interaction with the environment. In the last years, mobile devices have increased the number of embedded sensors they have. Accelerometers, cameras, gyroscopes, etc. enable more intuitive interfaces: pointing an object to get information about it or using a 2D sensor reader to make easier the information search process also increase the development of mobile context-awareness.

Table 1: A list of context-aware features.

<table>
<thead>
<tr>
<th>PERSONAL</th>
<th>CONTEXT-AWARE FEATURES</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Identity and profile management</td>
</tr>
<tr>
<td>2</td>
<td>Preferences and group membership management</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>PHYSICAL</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
</tr>
<tr>
<td>4</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>ACTIVITY</th>
</tr>
</thead>
<tbody>
<tr>
<td>5</td>
</tr>
<tr>
<td>6</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>RESOURCES MANAGEMENT &amp; COMMUNICATIONS</th>
</tr>
</thead>
<tbody>
<tr>
<td>7</td>
</tr>
<tr>
<td>8</td>
</tr>
<tr>
<td>9</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>INTERFACES</th>
</tr>
</thead>
<tbody>
<tr>
<td>10</td>
</tr>
</tbody>
</table>

Elaborating on these questions, Table 1 gathers some features enabling context-awareness. From their combination, mobile services may be built: features 1, 2, 3, 5 and 7 may be identified in a mobile social software application, while a sport monitoring service may need to handle characteristics 1, 3, 4 and 6.

In the next Section we analyze the implementation of these features in the stored data.

5.2 Some Results

- Limited but detectable use of context-aware features. 20% over the total of analyzed services are using personal descriptors, only 8% are employing physical parameters and managing resources and communications, while 5% of the services are using activity descriptors. At least one of the considered features appears in 47% of the 735 services. Around 49% of these services are just implementing one of the features; in fact, only 9.4% employ more than four context-aware characteristics.

- Upward trend in personalization and emergence of sensor-assisted HCIs. Discarding the incomplete series of 2003 and 2008 and aggregating the rest of the services in two biannual periods (2004-2005 and 2006-2007), Figure 4 shows a general increase of context-awareness.

Personalization (identity and profile, preferences and group membership) appears as a general trend. Real time information about the user connectivity (logical state) is also increasingly used. On the other hand, peer to peer communications have crossed the barrier of mobility in the second period of time. Almost 16% of the new mobile applications also manage location descriptors. Finally, services using new interfaces have significantly augmented in the period 2006-2007.

Figure 4: A biannual comparison of context-awareness implementation in mobile services. Percentage over the total production in the periods of analysis (300 services in 2004-2005 and 304 services in 2006-2007).

- LBS lead the group of precontext-aware services, together with information,
communications, personal information managers and social networks applications. Figure 5 compare the implementation of different context-aware features based on the categories presented in Section 4.1. It shows that Location and tracking services are the most personalized and activity aware ones. Personal information managers do not lag behind with respect to these two features, although in general they do not include location as a descriptor. On the contrary, information services are increasingly using positioning as a filter. Mobile commerce services do not use activity data but are personalized and sometimes related to the user location. In general, “entertainment and self content generation” services are scarcely considering the use of location.

![Figure 5: Number of services for each category that implement personal, activity, physical and resource related features.](image)

![Figure 6: Evolution of context-awareness in the periods 2004-2005 and 2006-2007. Data for some of the categories identified.](image)

6 CONCLUSIONS

This work explains the results of the analysis of a large database of mobile services, systematically updated during the period 2003-2008. Apart from giving a general view of mobile services evolution, we have focused on studying to which extent context-awareness is implemented in commercial mobile services. With this purpose, we have identified several context-aware descriptors (personal, physical or activity related) and features (capability of managing resources, establishing P2P communications or implementing advanced HCIs), and evaluated their implementation in mobile services.

- In general, few commercial mobile services can be considered fully ‘context-aware’, although isolated features have been detected in about half of them.
- Personalization -this understood as the capability of handling identity, preferences and group membership information- is the context-aware feature that most applications implement. Physical and activity descriptors’ use lags behind.
- Mobile services are increasingly incorporating the use of context-awareness, although there is not a breaking point in the period 2004-2007. Again, personalization is the most significant trend.
- Context-aware features have been detected in services of all the functional categories. Anyway, services in the following categories “Location and Navigation”, “Social networks”, “Information” and “Personal information management” are showing the highest levels of context-descriptors’ use.
- Mobile ‘social networking’ applications may be considered as highly context-aware. They have remarkably appeared during 2006-2007 and are intensively using personal descriptors while discovering and managing resources. Location-based filtering is also included in some mobile social services.
- The integration of sensors such as gyroscopes or accelerometers is making possible new interaction mechanisms with the environment. Sensors assisted HCIs proposals have noticeably grown in 2006-2007.
Nowadays favourable boundary conditions - progressively more connected environments, evolved mobile devices (with new communication capabilities, more usable HClIs and embedded sensors), "literate" users and more established data markets – seem enough to make context-aware services enter in the commercial offer. This “context-awareization” of mobile services has already started, even if there is still a long way to go.

ACKNOWLEDGEMENTS

This work has been financed by the Spanish Ministry of Education and Science under grant TSI2005-07344 (COLOCAME) and by the Government of Madrid under grant S-0505/TIC-0255 (MADRINET).

REFERENCES


AN ACCESS-CONTROL MODEL FOR MOBILE COMPUTING WITH SPATIAL CONSTRAINTS
Location-aware Role-based Access Control with a Method for Consistency Checks

Michael Decker
Institute AIFB, University of Karlsruhe (TH), Englerstr. 11, 76 128 Karlsruhe, Germany
decker@aifb.uni-karlsruhe.de

Keywords: Mobile Information Systems, Access Control, Location-based Services, Mobile Security.

Abstract: Some of the most salient challenges that come along with the employment of mobile information systems stem from security issues: portable devices like PDAs, smartphones and notebooks easily get stolen or lost and wireless data transmission could be eavesdropped, so that unauthorized individuals gain access to confidential resources. One approach to tackle these problems is location-aware access control, i.e. based on knowledge about the user’s position the information system can decide if access to a resource should be granted or not. For example a nurse using a PDA should only be allowed to access confidential patient data while staying on the premises of the hospital. In our article we present a data model for location-aware access control based on the concepts of roles. Using our model it is possible to assign location restrictions to several entities, e.g. to users, to roles or permissions. We also propose a method to analyze the consistency of spatial constraints expressed by an instance of our model.

1 INTRODUCTION

Nowadays it is possible to build portable computers like PDAs and smartphones which are more powerful than stationary systems were a few years ago. At the same time technologies for wireless data communication like GPRS, WiFi or UMTS were developed and are nowadays widely available. Based on these two technologies mobile information systems (MIS) can be realized to make computer support available almost anywhere and anytime. These MIS have a great potential for many novel application scenarios, especially with regard to support mobile workers. But the utilization of mobile technologies comes along with serious security challenges, because due to their portability and small size mobile devices easily get stolen or lost. If unauthorized individuals get into the possession of a mobile device they can gain access to confidential resources (e.g. health records, telephone numbers of board members, financial data).

In our work we employ the concept of location-aware access control to address this problem field. The basic idea is that for the decision whether a particular user is allowed to access a given resource the current location of the mobile device is regarded. For example the access to the customer database using a mobile device should only be granted if the mobile user just stays on the premises of the company; downloading a confidential document to a notebook computer could be denied if the user stays abroad. To determine the location of a mobile device there are several technologies available (e.g. GPS or CellID), see Küpper (2005) or Hightower & Borriello (Hightower and Borriello, 2001) for an overview.

To implement location-aware access control a formalism is required to express which resources can be accessed at which location by which users. Such a formalism is called “access control model” (ACM). In our article we will introduce a location-aware ACM that is based on the notion of role-based access control (RBAC). Location-restrictions can be assigned to several elements in this model to provide a great degree of versatility. We also introduce methods to analyze model instances based on geometric operations; as far as we are aware this concept is novel and not mentioned in pertinent literature yet.

The remainder of the article is structured as follows: We will first give a short introduction to access control models in section 2.1 before we introduce our own model and give an example scenario in section 2.2. A formal presentation of the model will be given in section 3; this comprehends also a simple location model. Based on this formalization we describe some ways to analyze model instances by geometric con-
We also implemented a small system to visualize the location constraints of model instances and to run the analyzing algorithms; this implementation will be sketched in section 5 before we conclude in section 6.

2 ACCESS CONTROL MODELS

2.1 Basics

Access Control Models (ACM) are formal models to describe under which conditions an information system should grant access to resources. There are three main groups of ACM (Samarati and di Vimercati, 2001): mandatory access control (MAC), discretionary access control (DAC) and role-based access control (RBAC).

Using MAC systems subjects as well as objects are classified according to their level of confidentiality (e.g. "top secret", "secret", "confidential", "unclassified"). Then there are rules that impose restrictions like "a subject should only be allowed to read an object if the subject’s classification (or clearance) is at least as high as the one of the object." One prominent example for such MAC is the Bell-LaPadula-model.

DAC models follow the owner principle: each object under protection of the access control model belongs to an owner. This owner can grant rights to perform certain operations on that object to other subjects. Examples for DAC-models are those used by operating systems, e.g. on Unix a user can grant the right to read a file he just created to users of a so called group. For the article at hand the most important group of ACM is RBAC (Ferraiolo et al., 2003). The basic notion is to assign permissions not to individual subjects but rather to so called roles. Roles represent job descriptions or positions within an organization. So if Alice is hired as new "manager" the permissions required for this job (e.g. read payroll file, approve order) don’t have to be assigned individually; we just have to assign the "manager" role to Alice’s user account.

Almost all location-aware access control models that can be found in literature are extensions of RBAC. The basic idea to obtain a location-aware RBAC model is to switch particular elements of the RBAC on or off depending on the user’s location, e.g. roles (Bertino et al., 2005), the role-permission-assignment (Hansen and Oleshchuk, 2003) or the user-role-assignment (Chandran and Joshi, 2005).

2.2 A Location-Aware Access Control Model based on RBAC

In this subsection we will introduce an RBAC-based model for location-aware access control which is depicted in figure 1 as entity-relationship diagram. The cardinalities in the diagram are given in the form of \((min, max)\)-values. For example \((1, *)\) as cardinality of entity "role" in the "role-location"-relationship means that each role entity occurs at least once in the relation representing that assignment. Since \(* = max\) a role entity may occur as many times as needed in that relation.

Each user can be assigned to several roles. Roles again are collections of permissions. Permissions in RBAC-models are usually considered as symbols that have to be interpreted for the respective deployment. In the figure we interpret a permission as a set of operations (e.g. read, write, update) that can be performed on a given class of objects. An object class might represent a class of documents like "customer record". For each object class there might be several object instances, e.g. "customer record for Mr. Meyer" and "customer record for Ms. Miller". The entity location represents a location like a city, a room or a country. We will later give a formal definition for what a location is.

Several components of the model have a relation to at least one location. A component is only enabled if the respective user stays at that location, e.g. if the role "manager" is restricted to the locations "local branch Munich" and "local branch Lisbon" this means the role can only be used when the user is in Munich or Lisbon. If there should be no location restriction we just assign the location "universe" (which covers the whole reference space) to the respective component. The remaining points for the assignment of location restrictions are:

- Restricting a user to locations means that the user can use the system only at certain locations, e.g. we can restrict a user to "southern Spain" because he has only to work in this region for the organization or he bought a software licence that covers only this part of the country.
- Restricting a permission to locations means that this permission should only be accessible if the user is at the specified locations, no matter how powerful his role is. An example is that "access payroll" should only be allowed while being on the company’s premises.
- It is also possible to assign location restrictions for individual objects, e.g. the object "research report no. 123" is not allowed to be accessed outside a...
A certain country because espionage is feared.

- Attaching a location restriction to the user-role-assignment means that the given user is allowed to play that role only at certain locations, e.g. a mobile worker is only allowed to play the role “travelling salesman” in his personal sales district.

- Attaching location restrictions to the role-permission-assignment means that the given permission can be only performed using that role when the user is at certain locations, e.g. we could have a role “manager” that encompasses the permission “approve order” but only within Europe.

We don’t assume that for a real-world instance of our model all six location restrictions will be used; but having six possible components where location-restrictions can be assigned to gives the freedom to choose the points that are most appropriate or natural for the respective scenario. If location restrictions are assigned to more than one point in the model it might be necessary to calculate the intersection of different location restrictions to obtain the area where the user is indeed authorized to perform a certain permission.

To exemplify our model we give the following example: The scenario encompasses a company with three local branches that sends mobile employees to facilities in a given area (e.g. to deliver goods or to perform some kind of maintenance work). In figure 2 the location restrictions for four roles (dotted lines) and three permissions (solid lines) are depicted as rectangles: \( p_1 \) is the permission to access the payroll file; since this is confidential information this should only be possible while staying at one of the three local branches. \( p_2 \) is the permission to access customer data; this permission is restricted to three service districts where the facilities are that have to be served by the company’s employees. \( p_3 \) is a navigation service licensed from an external service provider; this service is accessible only in an area that covers most of the three services districts (licensing for a larger area would be more expensive).

Roles \( r_1 \) to \( r_3 \) are roles for mobile workers (e.g. service technicians): \( r_1 \) is for the northern part of district 1, \( r_2 \) for the southern part of that district. The number of jobs to perform in districts 2 and 3 isn’t big, so a single role \( r_3 \) is sufficient for both. \( r_4 \) is the manager’s role who has to visit facilities in all districts, e.g. to handle complaints or to acquire new customers. Therefore this role’s location constraints encompass all districts. Role \( r_3 \) has the permissions \( p_2 \) and \( p_3 \); this means that a user having \( r_3 \) can access customer data (\( p_2 \)) in the areas where the rectangle of \( r_3 \) intersects with the two rectangles for permission \( p_2 \); the navigation service is only accessible where the rectangle for \( r_3 \) intersects with the rectangle for \( p_3 \).

3 FORMAL DESCRIPTION OF THE MODEL

3.1 Location Model

To give a formal description of our model and to describe spatial analyzes to be performed on instances of the model we need a formalization of the concept of a “location”:

We denote the area to be covered by the ACM as reference space or *universe*. Depending on the scope...
of the MIS universe may be the premises of a company, a country or earth’s surface. All other locations like cities, regions, buildings or rooms are real subsets of the universe: \( l \subseteq \text{universe} \). The set of all locations \( l \) and the universe is named \( \text{locs} \):

\[
\text{locs} = \{ l | l \subseteq \text{universe} \}
\]

With regard to the later implementation we demand that all locations are polygons because locations will be stored in a database system. Since we can approximate every area by a polygon with arbitrary precision this is a weak restriction. If \( PT \) denotes the location "Portugal" and \( LIS \) the location "LISBON" we get:

\[
\text{locs} = \{ \text{universe}, PT, LIS, \ldots \}
\]

For each location or set of locations we can calculate the covered area with the function \( \text{area}() \), which will return a real number greater than zero:

\[
\text{area}(): 2^{\text{locs}} \rightarrow \mathbb{R}_+
\]

If two locations of an input set overlap the intersecting area will be counted only once.

For two locations \( l_1, l_2 \in \text{locs} \) we can calculate the intersection and the union. When calculating the intersection of two polygons we may obtain one polygon, several polygons or the empty set as result, so the intersection operator’s range is the power set of \( \text{locs} \):

\[
l_1 \cap l_2 \rightarrow 2^{\text{locs}}
\]

The union set of \( l_1 \) and \( l_2 \) can consist of one location (if \( l_1 \) and \( l_2 \) intersect or touch each other) or two locations, but the empty set isn’t possible:

\[
l_1 \cup l_2 \rightarrow 2^{\text{locs}} \setminus \emptyset, \quad |l_1 \cup l_2| \in 1, 2
\]

There is also the set \( C \) of location classes. Location classes are semantic categories for locations, e.g. \( \text{C} = \{ \text{cities, countries, buildings, …, unclassified} \} \). Each location in \( \text{loc} \) is mapped to exactly one location class in \( C \):

\[
\text{class}(): \text{locs} \rightarrow C
\]

One class in \( C \) is "unclassified" which is the "dummy class” to be used if there is no reasonable semantic category for a location. Location classes are a mean to support human users when browsing through the set of pre-defined locations.

3.2 Core Model

We have a set for each of the main entities of the ACM: the set of users \( U = \{ u_1, u_2, \ldots \} \), the set of roles \( R = \{ r_1, r_2, \ldots \} \) and the set of permissions \( P = \{ p_1, p_2, \ldots \} \). For the sake of brevity we don’t consider the classes and objects behind a permission. Further there are the sets \( UR \) and \( RP \): if user \( u_0 \) has role \( r_0 \) then \( (u_0, r_0) \in UR \) and if permission \( p_0 \) is assigned to role \( r_0 \) then \( (r_0, p_0) \in RP \). Together these sets form the set of main entities \( E \):

\[
E = U \cup R \cup P \cup UR \cup RP
\]

The function \( \text{loc}() \) takes as input one of these entities and returns the subset of locations the respective entity is restricted to:

\[
\text{loc}(): E \rightarrow 2^{\text{locs}} \setminus \emptyset
\]

The locations returned by \( \text{loc}() \) are also denoted as a user’s, a role’s resp. a permission’s area. Example:

\[
\text{loc}(u_1) = \{ PT, MUN \}
\]

This means the user \( u_1 \) is allowed to use the system in Portugal or Munich. If an entity isn’t restricted to a location at all \( \text{loc}() \) just returns the universe. If we want to express that user \( u_2 \) can activate role \( r_2 \) only in Lisbon this would be:

\[
\text{loc}((u_2, r_2)) = \{ LIS \}
\]

4 SPATIAL ANALYSES OF MODEL INSTANCES

In this section we will describe some approaches to perform spatial analyses on instances of our model to find inconsistent configurations. For the sake of simplicity we don’t consider location restrictions assigned to user-role-assignments and role-permission-assignments.

4.1 Coverage

Coverage is a function that takes one entity from the set of all users, roles and permissions as first argument (pivot entity) and one of the target categories \( T = \{ \text{"user", "role", "permission"} \} \) as the second argument:

\[
\text{cover}(): (U \cup R \cup P) \times T \rightarrow 2^{\text{locs}}
\]

If the pivot entity is of the type specified by the target category then \( \text{cover}() \) just returns the location restriction of the pivot entity, e.g.\n
\[
\text{cover}(u_0, \text{"user"}) = \text{loc}(u_0)
\]

If the pivot entity stands "more left” than the target category in figure 1 then the result is the set of locations where the pivot entity can activate at least one entity of the target category:

- \( \text{cover}(u_0, \text{"perm"}) \) returns the area where user \( u_0 \) can activate at least one permission.
Here the location area of role \( r_0 \) at least one permission can be activated.

In the opposite case (the pivot element is on the right-hand side of the target category) the locations returned are the area where the pivot element can be activated by at least on entity of the target category:

- \( \text{cover}(r_0, \text{"perm"}) \) returns the area where with role \( r_0 \) at least one permission can be activated.
- \( \text{cover}(p_0, \text{"role"}) \) returns the area where permission \( p_0 \) can be activated by at least one role.
- \( \text{cover}(p_0, \text{"user"}) \) returns the area where permission \( p_0 \) can be activated by at least one user.
- \( \text{cover}(r_0, \text{"user"}) \) returns the area where role \( r_0 \) can be activated by at least one user.

Due to space limitations we will only discuss \( \text{cover}(p_0, \text{"user"}) \) in detail, which is also the most interesting case: if permission \( p_0 \) is location-restricted to certain locations (e.g. area where customers have to be served, area for which a software license was bought) we may want to check if every point of the permission area is covered by at least one user, i.e. is there a part of the permission area where no user according to his roles and his own location restriction can use that permission? If this case is given it would mean that we had customers that cannot be served or that we bought the software license for a too big area.

The formal definition is as follows:

\[
\text{cover}(p_0, \text{"user"}) = \text{loc}(p_0) \cap (\bigcup_{(r_0,p_0) \in \text{RP}} \text{cover}(r_0, \text{"user"}))
\]

So we have to calculate the intersection of the permission area with the union of all coverages for roles that are assigned to \( p_0 \). The coverage of a role again is:

\[
\text{cover}(r_0, \text{"user"}) = \text{loc}(r_0) \cap (\bigcup_{(u_0,r_0) \in \text{UR}} \text{cover}(u_0, \text{"user"}))
\]

Here the location area of role \( r_0 \) has to be intersected with the union of the area of all users that are assigned to role \( r_0 \).

If the permission area is smaller than the coverage, i.e.

\[
\text{area}(\text{cover}(p_0, \text{"user"})) < \text{area}(\text{loc}(p_0))
\]

we have locations where the permission could be performed, but no user is available who could. This is a strong hint for an erroneous configuration.

To exemplify this we give an example with two roles and three users. For the user-role-assignment we have:

\[
\text{UR} = \{(u_1, r_1), (u_2, r_1), (u_2, r_2), (u_3, r_2)\}
\]

Permission \( p_0 \) is assigned to both roles. The respective location restrictions are depicted in figure 3. The coverage is smaller than the location restriction assigned to \( p_0 \) because there is an uncovered area (the shaded part of the uppermost permission-area). This part is covered by role \( r_2 \). User \( u_2 \) and \( u_3 \) are assigned to that role, but their location restrictions don’t cover the shaded area.

### 4.2 Empty Assignments

We talk about an “empty user-role assignment” when for \((u_0, r_0) \in \text{UR}\) the following holds:

\[
\text{loc}(u_0) \cap \text{loc}(r_0) = \emptyset
\]

The intersection between the locations assigned to the user and the role is empty, so the respective assignment is redundant, i.e. it could be removed without changing the policy.

Further the following case shouldn’t occur:

\[
\text{loc}(u_0) \cap \text{cover}(r_0, \text{"perm"}) = \emptyset
\]

This means that role \( r_0 \) was granted to user \( u_0 \), but this assignment doesn’t allow him to perform any permission. In the same way we can check for “empty role-permission assignment” for \((r_0, p_0) \in \text{RP}\):

\[
\text{loc}(r_0) \cap \text{loc}(p_0) = \emptyset
\]

Each assignment made by the administrator of a model should be checked for these two cases; if detected the administrator should be asked if he is sure about this assignment.

### 5 IMPLEMENTATION

We developed an application with a Java/Swing-based graphical user interface to have a runtime environment for instances of the proposed ACM. The application can perform the spatial analyses covered in the last section and is also capable of visualizing location constraints assigned to individual entities.

The instances of the ACM are stored in a PostgreSQL/PostGIS database management system which provides support for working with spatial data. For the visualization of spatial data we resorted to OpenJump.

OpenJump can load spatial data from a database into different layers. We use this feature to load the location areas for different entities of a model instance into various layers to see how they are related. Figure 4 is a screenshot from the module of the application that implements the coverage calculation as described in section 4.1. On the right side we can see the OpenJump-GUI that displays the covered and the uncovered area of the example given in figure 3.
6 CONCLUSIONS

We started by motivating the need for location-aware access control and proposed a RBAC-based model where location-restrictions can be assigned at several components to provide a great degree of flexibility. Based on a formalization of the model itself and the underlying location model we introduced some approaches to analyze model instances based on spatial calculations; this helps to detect erroneous location restrictions. We also sketched the prototypical implementation of a system as runtime environment for the model and visualization purposes.

REFERENCES


A PLATFORM FOR INVESTIGATING EFFECTIVENESS FOR STATIC, ADAPTABLE, ADAPTIVE, AND MIXED-INITIATIVE ENVIRONMENTS IN E-COMMERCE

Khalid Al-Omar and Dimitris Rigas
Department of Computing, University of Bradford, Bradford BD7 1DP, U.K.
K.Al-Omar1@Bradford.ac.uk, D.Rigas@@Bradford.ac.uk

Keywords: Adaptive, Adaptable, Mixed-Initiative, Static, Usability, Effectiveness, Interactive.

Abstract: This paper introduces an empirical study to investigate the use of four interaction conditions: Static, Adaptable, Adaptive, and Mixed-initiative. The aim of this study is to compare the effectiveness of these four conditions with regard to the number of tasks completed by all users and the number of users who completed all tasks. In order to carry out this comparative investigation, four experimental interfaces were built separately. These environments were tested independently by four separate groups of users, each group consisting of 15 users. The results demonstrated that in the searching tasks the most effective condition was the Mixed-Initiative. In the learnable tasks the most effective condition was the Adaptable condition. In addition, the Static approach was found to be less effective than all other approaches.

1 INTRODUCTION

Today, software application and e-commerce web-based application (Alotaibi and Alzahrani, 2004) is crowded with functions, icons, menus, and toolbars (McGrenere et al., 2007). In addition, the web-based e-commerce application is crowded in both the Graphical User Interface and content. This is a phenomenon called ‘Bloatware’ or ‘featurism creeping’ (McGrenere et al., 2007). This phenomenon makes searching for information and products within e-commerce web-based application very complex (Findlater and McGrenere, 2004) (Te’eni and Feldman, 2001). Therefore, personalising the application to users need and preferences is essential and becomes very important (Findlater and McGrenere, 2004), (Fink et al., 1998).

Personalisation is a topic of debate between two communities, the Intelligent User Interface community favouring adaptability (Shneiderman, B. and P. Maes, 1997) at the expense of user freedom and Human Computer Interaction community favoured adaptability (Shneiderman, B. and P. Maes, 1997) at the expense of system help. According to McGrenere et al. (2002) there are three potential ways to personalisation: 1) by users and this is called an adaptable approach. 2) by system and this is called An adaptive approach. 3) by both the users and system and this is called Mixed-initiative approach which is a combination of adaptable and adaptive approach.

Despite the disagreement in the research community, there are multiple direct comparisons between Static, Adaptable, and Adaptive approaches have shown different results. In 1985, the first study of adaptation was reported by Greenberg and Witten. They demonstrated an adaptive interface for a menu-driven application. In their study users were novices on the task and the interface (Greenberg and Witten, 1985). In addition, Greenberg and Witten (1985) built a directory of telephone numbers that users can access through a hierarchy of menus. Their goal is to reduce the number of key-presses buttons. Their approach is to present items at a level in the hierarchy according to the number of selection. Greenberg and Witten tested their system against a static system in a 26-participant experiment. Their results showed that subjects performed faster with the adaptive system, and 69% of subjects prefer the adaptive system. In addition, they found that the adaptive system reduces the search paths for repeated names, reduce 35% in time per selection, and reduce 40% in errors per menu. Trevellyan and Browne (1987) replicated the Greenberg and Witten’s experiment with a larger number of trails because they believe after a large of trails subjects
will be familiar with the static and they can memorised the sequence of key-presses. This would reduce the mean time per menu. However, they found that the adaptive system is effective and after using the system for long period of time users did begin to perform better with the static interface. This study did not provide a firm conclusion since the total number of subjects in each interface is 4 subjects.

In 1989, Jeffrey Mitchell and Ben Shneiderman (1989) conducted an experiment to compare an adaptable interface that items positions change dynamically according to frequently clicked item, with a static menu. Thirty-three subjects assigned randomly tried both menus and carried out the same 12 tasks in each menu. Their results showed that static menu faster than the adaptable menu at first group of tasks, and no difference in the second group of tasks. That because, subjects in both groups were able to increase their performance significantly. However, Eighty one percent of the subjects preferred the static menu. Another study introduces a system to provide environment for adapting Excel’s interface to particular users (Thomas and Krogsæter, 1993). The result showed that an adaptive component which suggests potentially beneficial adaptations to the user could motivate users to adapt their interface. Jameson and Shwarzkopf (2000) conducted a laboratory experiment with 18 participants a direct comparison between automatic recommendations, controlled updating of recommendations, and no recommendations available. Their comparison concerned about the content rather than the Graphical User Interface. Their results showed that there was no difference on performance score between the three conditions.

In 2002 McGrenere et al. conducted a six-week with a 20 participant field study to evaluate their two interfaces combined together with the adaptive menus in the commercial word processor Microsoft Word 2000. The two interfaces are a personalised interface containing desired features only and a default interface with all the features only. The first four weeks of the study participants used the adaptable interface, then the remaining for the adaptive interface. 65% of participants prefer the adaptable interface and 15% favouring the adaptive interface. The remaining 20% favouring the MsWord 2000 interface. This work extends by Findlater and McGrenere (2004) and they compared between the static, adaptable, and adaptive menus. Their result concludes that the static menu was faster than the adaptive menu and the adaptable menu was not slower than the static menu. In addition, it shows that the adaptable menu was preferable than the static menu and the static was not preferable to adaptive menu. Another study examined how characteristics of the users’ tasks and customisation behaviour affect their performance on those tasks (Bunt et al., 2004). The results confirm that users may not always be able to customise efficiently. The results indicate that customisation is beneficial to reduce tasks time if it done right. Also, indicate that the potential for adaptive support to help users to overcome their difficulties.

In 2005, Tsandilas and Shraefel conducted an empirical study that examined the performance of two adaptation techniques that suggest items in adaptive lists. They compared between the baseline where suggested menu items were highlighted and shrinking interface which reduced the font size of non-suggested elements. The results indicate that the Shrinking information was shown to delay the searching of items that had not been suggested by the system. In addition, the accuracy affected the ability of participants to locate items that were correctly suggested by the system. Gajos et al. (2005) comparing two adaptive interfaces: 1) their Split interface, which is most of the calculator’s functionality was placed in a two-level menu. 2) Altered Prominence interface, all functionality was available at the top level of the interface. The study showed user preference for the split interface over the non-adaptive baseline. Another experiment compared the learning performance of static versus dynamic media among a 129 students. Their result showed that the dynamic media (animation lessons) has a high learning performance than the static media (textbook lessons) (Holzinger, 2008).

Despite the debate between the two communities, there has been very little work directly comparing to either an adaptive or adaptable approach with the Mixed-Initiative approach through empirical studies. On example of a such a comparison conducted by Debevec et al. (1996). They compared between their adaptive bars with the built-in toolbar present in MSWord. Their results showed that the mixed-initiative system improved significantly the performance in one of two experimental tasks. Bunt et al. (2007) designed and implemented the MICA (Mixed-Initiative Customisation Assistance) system. Their system provides users with an ability to customise their interfaces according to their needs, but also provides them with system-controlled adaptive support. Their results showed that users prefer the mixed-initiative support. Also, it shows that the MICA’s
recommendations improve time on task and decrease customisation time.

2 THE EXPERIMENTAL PLATFORM

An experimental e-commerce web-based platform was developed to be used as a basis for this empirical study. The platform provided four types of interaction conditions: Static, Adaptable, Adaptive, and Mixed-initiative. The structure of the platform is similar to many e-commerce web-based platforms. The difference between the four conditions applied to the contents, layout, and item position on the list.

2.1 The Static Platform

The layout, content, and item position on the list does not change during the course of usage. Our goal was to design the ideal platform to do the required tasks as efficiently as possible.

2.2 The Adaptive Platform

The layout, content, and item position on the list does change by system during the course of usage. Adaptation helped users to find items by changing content to their preferences. Our goal was to design the most predictable personalised approach as possible.

2.3 The Adaptable Platform

The layout, content, and item position on the list is changed by the user during the course of usage. Our goal was to make the customisation process as easy as possible. Therefore, the Coarse-grained and Fine-grained (Findlater and McGrenere, 2004) customisation techniques were utilised by allowing the user to move items to a specific location (See Figure 2). However, the main page provides two choices for the user to choose from. The first choice is an empty page that is left to the user’s decision as to which content to add in. The second choice is full content that has already been suggested. This is because some of the early studies suggested a need to examine full-featured interfaces versus reduced interfaces. However, when the participant started, four items were displayed as a default in each web part of the home page. Subjects can increase the number of displayed items as many items as they like and reduce the number of displayed items not less than one item. In addition, subjects can sort the web contents by item name, id and price and the user can also search in different sub-categories. Subjects can add new content to the home page, delete, and move an existing content to different positions.

Therefore, the adaptive approach algorithm dynamically determines item position on the list based on the most frequently and recently used items. The two algorithms are used by Microsoft (Findlater and McGrenere, 2004) and suggested by the literature (Findlater and McGrenere, 2004). For our experiment, once the user clicks the items they will move up to the top of the list (See Figure 1).
2.4 The Mixed-Initiative Platform

In the Mixed-Initiative condition the control is shared. Therefore, our goal was to make sure the control is shared as fairly as possible. The Mixed-Initiative condition algorithm is dynamically determined based on the most frequently and recently used items. However, to allow users to take control, a new function was implemented to lock and unlock item movement (See Figure 2). Items will be moved up to the top of the list when clicked three times, even if the list locked. Initially, when the website is loaded the default content of the home page is personalised. However, organising the list is the user’s responsibility along with locking the lists.

3 EXPERIMENTAL DESIGN

The experimental platform was tested empirically by four independent groups, consisting of 15 users. All the groups of users were asked to accomplish the same 12 tasks. These tasks were designed with three complexity levels: easy, medium, and difficult. In order to avoid the learning effect, the order of the task complexity was varied between subjects. The number of available items, item position (location) in the list, number of requirements and guidance was considered when designing the tasks, i.e. more than three items available within a list that consists of a maximum of 20 items. The items are positioned at the top, middle and at the end of the list. Thus users can find the item even if the list changes. The number of requirements is less than four. The users are guided to the list by providing the name of the list and the subcategory.

<table>
<thead>
<tr>
<th>Category</th>
<th>Easy Tasks</th>
<th>Medium Tasks</th>
<th>Complex Tasks</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of requirements</td>
<td>1-3</td>
<td>4-5</td>
<td>6-10</td>
</tr>
<tr>
<td>Number of clicks required</td>
<td>1 to 2 clicks</td>
<td>5 to 10 clicks</td>
<td>More than 15 clicks</td>
</tr>
<tr>
<td>Number of visited pages required</td>
<td>0 pages</td>
<td>1-2 pages</td>
<td>More than 5 pages</td>
</tr>
<tr>
<td>List size</td>
<td>10 to 15 items</td>
<td>25 to 35 items</td>
<td>35 to 50 items</td>
</tr>
<tr>
<td>Items availability</td>
<td>3 to more</td>
<td>1-3</td>
<td>1</td>
</tr>
</tbody>
</table>

Table 1: Tasks design.

For the medium tasks, the number of available items is reduced to two items within a list that consists of more than 30 items. The items are positioned at the middle of the list. The number of requirements is more than four and up to six requirements. The users are guided to the list but not the subcategory, so it is the user’s responsibility to search for items in the subcategory.

For the difficult tasks the number of available items is one item within a list that consists of more than 40 items. The items are positioned at the middle of the list, to make sure that users can find the item even if the list changed. The number of requirements is more than seven. In the difficult tasks there is no guidance to items, so it is the user’s responsibility to search for items in all lists and all subcategories.

4 SUBJECTS

These environments were tested empirically by four independent groups, each group consisting of 15 users. All the groups were asked to accomplish the same group of tasks (three easy tasks, three medium tasks, and three difficult tasks) and a one learnable task before starting each group. Each user attended a five minute training session about their environment before doing the requested tasks. A pre-questionnaire conducted before the experiments to obtain users personal information. All users were between the ages of 18 and 40. 44 of them were male, while the remaining 16 were female. 70% of them were postgraduate students. Most of the participants used the internet for 10 hours or more a week. 85% stated that they do not customise new software unless they have to; the remaining 15 stated that they do so. Also, 32% never used any customisable web pages, where 17% used it once, and just four participants used it every time they went online.

5 RESULTS AND DISCUSSION

Effectiveness was measured by calculating the percentage of users who completed (learning and completion) tasks along with the percentage of (learning and completion) tasks completed by all users. To compare the effectiveness between the four conditions, three critical time limits for task completion was derived for each level of tasks (easy, medium, and complex). Therefore, a task would be regarded as successfully completed if users completed the task within the level critical completion time.

However, it was noticed that during the experiment users who participated in the evaluation
of the adaptable and mixed-initiatives were more confident than the static and adaptive. Also, users got confused while participating in the evaluation of the adaptive and static conditions. This confusion made them spend time on understanding what is happening around them. Overall, just 8 users did not complete all tasks using the Mixed-Initiative whereas 23 users did not complete all tasks using the adaptive condition. In the adaptable condition, 24 users did not complete the all tasks whereas only 2 users did complete all tasks using the Static condition. This shows that the overall number of users who completed all tasks in the Mixed-Initiative is higher than the other conditions. An ANOVA result showed a significant difference in the number of users who completed the tasks at 0.05 (F = (3, 11), p <0.004). The users who completed the easy, medium and complex tasks using the Mixed-initiative condition is higher than the other conditions (Static, Adaptive, and Adaptable), excluding the users who completed the medium tasks using the Adaptive condition.

Overall, t-test was used to find out the difference between the four conditions. t-Test results showed that there was a significant difference of 0.05 between the number of users who completed all tasks using the Mixed-initiative condition compared to the adaptable (t(3)=4.38, cv=3.1) and static (t(3)=11.3, cv=3.1) conditions, but nothing significant was found when compared to adaptive (t(3)=2.04, cv=3.1). The users who completed the tasks using the adaptable and adaptive conditions are higher than the static condition. Also, it was found that the adaptable are higher than the adaptive in easy tasks and lower in medium tasks. Furthermore, there was a significant difference between the numbers of users who completed all tasks between the adaptable and static conditions (t (3) = 3.04, cv=3.1) and between the adaptive and static conditions (t (4) = 4.5, cv=2.7). Figure 1 shows the percentage of tasks completed by all users in each of the four conditions. However, the number of the tasks completed by all users was calculated to obtain an overall percentage. The result showed that the number of tasks not completed by all users was 8 tasks by using the Mixed-Initiative, 33 tasks by using the Adaptive, 44 tasks by using the Adaptable, and 83 tasks were not completed by using the Static.

In the learnable tasks, there was a difference between the four conditions (See Figure 4). This difference was found to be statistically significant at 0.05 by using the ANOVA test. T-Test results showed that there was a significant difference at 0.05 between the number of tasks completed by all users using the Mixed-initiative condition, compared to the Static condition (t(3)=11.3, cv=3.1) but not to the adaptive (t(2) = 2.6, cv = 4.3) and adaptable conditions (t(2)=3.1, cv=4.3). In addition, there was a significant difference between the Adaptive and Static conditions (t(4) = 4.5, cv=2.7). However, the number of users who completed all learnable tasks by using the adaptable condition was 11, which was higher than the other conditions. Following this was the mixed-initiative where 9 users completed their all learnable tasks, and the Static condition (3 users). The users who completed all tasks using the adaptive condition were lower (2 users) than all other conditions. The percentage of users who completed all tasks using the mixed-initiative condition was higher than the adaptive and static
conditions but not higher than the adaptable condition. The main reason behind this is that sometimes items’ positions in the lists changed without users’ noticing which caused them confusion.

Figure 4: Learnable Tasks.

6 CONCLUSIONS

This paper described an empirical study that was performed to investigate the effectiveness of the Adaptive, Static, Adaptable and Mixed-initiative conditions. In this investigation, the aim was to assess the effectiveness of these four conditions. One of the more significant findings to emerge from this study is that Mixed-Initiative approach was the best in terms of effectiveness in the searching tasks but not with the learnable tasks. In the learnable tasks the adaptable was better than all other approaches. In addition, the Static and adaptive conditions were found to be less effective than the other conditions in terms of number of tasks completed by all users and number of users who completed all tasks. Further work needs to be done to establish whether the presence and absence of multimodal metaphors on the mixed-initiative approach will help to make the most of the adaptive and adaptable advantages, at the same time as reducing their disadvantages.

REFERENCES

AN ENHANCED SERVICE PROVIDER COMMUNICATION INTERFACE WITH CLIENT PRIORITIZATION

Case Study on Fast-food Chain Restaurants

Slobodan Lukovic
ALaRI, University of Lugano, Lugano, Switzerland
slobodan.lukovic@lu.unisi.ch

Nikola Puzovic
Department of Information Engineering, University of Siena, Siena, Italy
puzovic@di.unisi.it

Milos Stanisavljevic
Microelectronic Systems Laboratory, EPFL, Lausanne, Switzerland
milos.stanisavljevic@epfl.ch

Keywords: e-Commerce, client prioritization, mobile devices, workload scheduling.

Abstract: With the increased dynamics of modern life, the efficiency and reliability of everyday services is emerging to be a fundamental concern. On the other hand, modern telecommunication technologies, like wireless Internet access, are penetrating all segments of our life. However, many every day activities and services still do not fully exploit new technologies. We propose an approach that enables increased deployment of E-commerce concepts in the fields where their usage was either small or negligible. Moreover, in the scope of the same concept, we introduce prioritization of clients in services where it was not commonly present to date. A solution for enhanced communication interface between service provider and customers is developed. As a case study, the system is designed and optimized for an implementation in a fast-food chain. The proposed solution is aiming at increasing of quality of service for customers, and at the same time increasing the operational efficiency of the provider. The main idea behind this approach is to enable customers to use their mobile devices, such as cell phones or PDAs, for browsing offered services or goods, viewing current service conditions and placing orders. We will detail theoretical concepts underneath and describe the implementation on both server and client side.

1 INTRODUCTION

Millions of people daily face different kinds of problems in common situations. These issues range from unpredictable traffic and parking problems to big waiting times in restaurants due to inefficient order/payment service. The rapidly growing urban population additionally increases pressure on service providers. We look for a response for these problems by incorporating e-Commerce concepts into service provider - client chain.

The basic idea of service order/payment process automation relies on the rapid increase of number of portable devices that are able to access Internet, and on the growth of the popularity of e-payment methods. It is estimated that in the next two years all manufactured cell-phones will be equipped with WiFi modules which will boost the availability of Internet services. Other broadband wireless services like UMTS and WiMax are becoming more widely available and affordable and the number of mobile users that will use these services will increase. Moreover, the number of online payments is constantly increasing and this way of payments is expected to be widely accepted in very near future. Putting all together, integration of all aforementioned services is logical consequence of technology development and of the evolution of users’ habits.

In this paper we detail, based on case study on fast food chain, a possible solution for incorporation and automation of the service scanning, ordering, payment and execution that increases efficiency, cutting the costs and bringing many other benefits for all sides involved in the process. Our aims are the improving of existing Quality of Services (QoS) in fast food restaurants, and at the same
time providing greater flexibility for both the service provider and consumers. We propose a fully integrated system that incorporates many different technological aspects, ranging from Internet browsing and e-payments to service performance evaluation and workflow scheduling. The final result is a novel communication interface that brings many new features and benefits to the client, like better service overview, waiting time prediction (service availability overview), prioritizing of clients, multilingual support, etc. On the other side, it cuts costs for service providers in terms of staff reducing, better insight in demand and market overview coupled with market profiling and targeted marketing. For now no experimental results are provided.

The paper is organized as follows: In Section 2 we summarize current state of technology and different types of its usage that are in scope of interest of this work, in Section 3 we present an overview of the overall architecture of the system. Algorithms and techniques used in realization of proposed concepts are detailed in Section 4. Finally, Section 5 presents conclusions and future work.

2 STATE OF THE ART

In parallel to development of communication technologies many different service providers have been adopting them to facilitate the interaction with clients, to provide new services or increase operational efficiency. On the other side mobile technologies such as WiFi, UMTS or WiMax as well as web access standards and protocols are more and more oriented toward better support for increased need for mobile Internet availability. In that sense many mobile web standards such as Wireless Markup Language (WML), Extended Modeling Language (XML) or XHTML etc. have been developed. Moreover, the development of AJAX (AJAX, 2008) has enabled exchange of small amounts of data between client and server, hence increasing the interactivity, speed and usability without the need to reload the entire contents of web page. This is especially valuable when devices have scarce computational and communication capabilities. By having more and more mobile devices online many new concepts got enabled. This especially concerns social networking (Eagle and Pentland, 2005), mobile commerce (Varshney et al., 2000) or intelligent wireless web (Alessio and Smith, 2001).

The most recent world-wide trend regarding wireless Internet access is deploying of WiFi Internet access at variety of locations such as airports, hotels, restaurants and so on (Friedman and Parkes, 2003); in some cases even free of charge (Smithers, 2007). At the same time online payment methods are getting widespread across wide range of activities (Weiner, 2000),(Ghosh and Li, 2007). Coffees and restaurants have made steps towards exploitation of WiFi at local service points (Friedman and Parkes, 2003) and some fast-food chains started using touch-screens deployed at tables (eTable) for offer browsing and order placing (VanLeeuwen, 2005).

We propose a novel solution for prioritization of clients using and enhancing different experiences and implementing them in fields where until now these technologies have not been commonly used. In order to provide better quality of service we model prioritized orders execution and service provider capacity with a well know operating systems task scheduling. The theoretical concepts adopted for purposes of our work will be discussed in detail in Section 4.

3 SYSTEM OVERVIEW AND IMPLEMENTATION CONCERNS

We envision a system that integrates WLAN access, priority order scheduling based on demand prediction and delivery automation that provides user-friendly interface to the client. For the purpose of the client-server communication we propose WiFi wireless Internet access technology since it is widely used and already deployed at many fast food points around the world. WiFi access represent an optimal compromise between simplicity and efficiency. The coverage, throughput and level of security are considerably greater than in the case of Bluetooth. On the other side it is very easy to deploy and cheap to use (from clients’ point of view it is free) in comparison to other broadband services as UMTS and WiMax. The range of the WiFi AP also makes it perfectly suitable for use in a local service point (a restaurant in this case). Nevertheless, the system is conceived in such way that it can be easily ported to other communication mediums. By using secure communication and already established methods of electronic payment, we will also provide high level of security that is necessary.

Figure 1 shows the overall organization of the system. The main components are servers in service points, central data warehouse and a connection to e-payment servers. The local servers are responsible for processing clients’ orders and updating central data warehouse. Payments are performed using secure connection to e-payment servers or by using in-house e-payment system. In the following subsec-
tions and we will also discuss the system from the participants’ point of view (both clients’ and service providers’), we will discuss the technology involved together with different implementation aspects.

3.1 Client’s View of the System

Client devices are PDAs or smart phones that are capable of running a web browser. User interface is implemented in HTML, and it is accessible from each web browser, so there is no need to install a special purpose application. Figure 2 shows the UML sequence diagram of the users’ interaction with the system. At the beginning, the client chooses appropriate WiFi Access Point (AP) that will consent him to access the services provided in the restaurant. Upon the entry to the network, clients are redirected to the start page from where they can browse the available services and current conditions of the system. There is a possibility to supply the client with additional information that traditional ordering methods cannot support (i.e. expected waiting time, multilingual interface etc.). The calculated expected waiting time coupled with delivery automation brings a possibility of introducing prioritization in the fast food service. For more details see Section 4.

From the start page clients will continue to the menu with products that are available in the restaurant. Through simple web forms clients can make a choice, and communicate the selection to the server in the service point when the order is completed. The system sends back the information about the order and order number. The payment can be done through SSL secured connection using credit-card payment, pay-pal and other methods of e-payment. In this case, a request for payment is communicated to the payment gateway (paypal server or bank server in case of credit card payment) that processes it and gives the confirmation. Another option is to use vendor provided vouchers that can be issued in the form of fidelity card, and can be used exclusively for payments inside the system, and they function in form of deposit or credit.

Once the payment is completed and the delivery is ready, it is placed in a delivery slot and the client receives a code associated with the order. The client types the code using keyboard on the delivery slot that is labeled with appropriate order number or with clients name. These slots can be implemented as parts of rotating table, where each slot has a protective cover. Once the correct code is inserted, the service is considered to be completed. The time between order placement and code insertion is considered as ‘service time’. This information is taken and used as correction factor for statistical processing and for calculation of the expected time of servicing.

3.2 Service Provider’s View of the System

The system in a service point processes e-Commerce orders in parallel with traditional ones (Figure 3). Requests from both sides go to the same server and their execution is scheduled in order of submission. The local server that processes orders is responsible for handling client requests and for scheduling the delivery of orders. Scheduling and delivery are performed taking into account also the priority of the clients. The algorithm for calculating waiting times is applied each time a new client enters the restaurant and this information is communicated to the client.

The technology used in the service point is rather simple and cheap, and it involves a server computer, wireless AP and Internet connection that will be used for communication with the central server. The algorithms that are used are described in subsequent sections.
4 SERVICE DEMAND AND SERVICE TIME PREDICTION MODEL

In this section we present mathematical concepts proposed for calculation of expected average service time considering prioritized clients. The main contribution reflects in the exploration of scheduling mechanisms in order to meet the deadlines - guarantee service. We briefly expose possible service classification concepts and compare them. The chosen concept is described in detail.

4.1 Definition of the Server and Client

We consider the server - service execution unit that takes the order, executes it and performs the delivery of the requested service or product. The server is characterized as multiple parallel process execution unit with a given total capacity - service power. The clients could be classified in three groups with two levels of priority:

1. Traditional clients served at the service line in a traditional way (worker at the counter table is serving one client after another from the FIFO - first in, first out, queue), taking low priority (priority equal 0).
2. Regular e-Commerce clients, performing orders through wireless system, taking low priority.
3. Prioritized e-Commerce clients, performing priority service orders through wireless system, taking high priority (priority equal 1).

Initially a certain server capacity is given to the prioritized clients. This capacity is dynamically adjusted during service execution in order to guarantee projected waiting time for prioritized clients and therefore quality of service (QoS). We have adopted following naming convention to describe different issues related to times spent in different phases of the order processing process:

- Queuing time - time spent in the line/queue while waiting for the order placing.
- Service execution time - time required to execute the requested service by the server.
- Service waiting time - the time that passes from the moment of performing the order to the moment in which the order is completed.
- Total service time - the time from the moment the client has entered service point to the moment when the required services are obtained.

4.2 Classification of Services by Processing Power Requirements

Almost all orders differ from each other in type of services requested, its quantity and quality which results in different service processing power requirements. Any order can be further decomposed in set of ‘atomic services’. We assume ‘atomic service’ to be the simplest possible single order (i.e. an ice cream or coffee). In that sense every order can be seen as a composition of various ‘atomic orders’. The most precise way of modeling the orders would be representing them in form of number of atomic services requests. Unfortunately this method would introduce huge computational overhead and the entire system would be very complex to implement. For the sake of the simplicity and better efficiency but without losing generality we classify orders according to number of atomic services requested into five classes ranging from small to huge ones. Each of those order types are assigned certain evaluated average service time. This evaluation is constantly updated by newly obtained data from the system.

4.3 Service Time Prediction Model and Scheduling for a Guaranteed Execution

Here we describe empirical model for run-time calculation of service expected time. This time is provided to clients as additional information and this property of the system is one more original feature that enhances existing services. The service time is modeled as a statistical variable with two components:

1. Non-stochastic component measuring the necessary service time for non-prioritized and prioritized services with a given scheduling scheme.
2. Stochastic component that incorporates the additional service time related to the average number of prioritized clients that are arriving during the time slot defined with non-stochastic component. The additional service time is added into the scheduling scheme as a prioritized process and a total service time is acquired after the probe scheduling execution. This component is acquired by processing empirical data related to number of prioritized clients and their orders that are dynamically updated every day. During the phase of initial system deployment these data are unavailable and therefore a worst case estimation is taken instead of statistical component.

The order execution inside the server is scheduled according to non-preemptive task execution scheduling. We consider this scheme to be the most suitable in case of fast food restaurants as preemptive service would cause disorder in execution of currently executed services and also would require different working concept inside the server (i.e., specialization to atomic tasks execution and work division that could support this would result in inefficient use of working power).

In case that prioritized tasks are not schedulable within dedicated server capacity for prioritized tasks, the amount of dedicated server capacity is increased to the minimal value that guarantees schedulability of prioritized tasks. This has a consequence of increase in waiting time for non-prioritized clients. However, QoS for prioritized clients is guaranteed.

Any practical scheduling algorithm in multiprocessor system presents a trade-off between performance and computational complexity. However, in our case scheduling computation time is not an issue (because it can be in the range of seconds) and we can explore the more complex scheduling schemes. Scheduling could be regarded as soft real-time or even non real-time problem. The Earliest Deadline First (EDF) algorithm is the most widely studied scheduling algorithm for real-time systems (Balarin et al., 1998).

EDF is more efficient than many other scheduling algorithms, including the static Rate-Monotonic scheduling algorithm. However, when the processing server is overloaded (i.e., the combined requirements of pending tasks exceed the capabilities of the system) EDF performs poorly. Researchers have proposed several adaptive techniques for handling heavily loaded situations, but they require the detection of the overload condition. Least Laxity First (LLF) algorithm (Ramamritham and Stankovic, 1994) is non-preemptive and selects the task that has the lowest laxity (the maximum time that a task can wait before being executed; laxity = total service time - service execution time) among all the ready ones whenever a processing server becomes idle, and executes it to completion.

Lee et al. (Lee et al., 1994) presents a fuzzy scheduling algorithm. Their proposed algorithm uses task laxity and task criticality as system parameters and doesn’t consider fairness. Their simulation model contains small number of tasks on a uni-processing unit system and they did not consider system overloads.

Chen et al. (Chen et al., 2005) proposed a scheduling model and a related algorithm that is suitable for both uni-processing and multiprocessor servers. They provide a method to detect work overloading and try to balance load with task dispatching. We propose to use model presented in (Hamzeh et al., 2007) using a fuzzy interface engine. The model we propose has a slight modification considering that there are only two levels of priority defined as "high" and "low". As shown in Figure 4, the major factors considered in used approach to determine the scheduling are task priority, total service time, service execution time, and used server capacity time. The notion of laxity is used in the proposed approach to facilitate the computation.

In proposed algorithm as shown in Figure 5, a newly arrived task will be added to the input queue. This queue contains the remaining tasks from last cycle that has not yet been assigned.

Fuzzy scheduler processes each task separately, computes its run-time priority and sends it to task dispatcher’s priority queue. In a multiprocessor system, this queue offers tasks to dispatcher by their run-time priority order (as shown in Figure 6). Dispatcher offers a new task whenever one of the processing units of the system finishes its current task.
1. For each task in input queue
   (a) Feeds task’s run-time priority using fuzzy inference engine
2. While a server has a free processing unit
   (a) assign the task with highest run-time priority to the processing unit
3. Loop forever
   (a) If a processing unit event occurs endenumerate
      i. Go to 2
   (b) If scheduling event occurs
      i. Update tasks parameters
      ii. Go to 1

Figure 5: Proposed algorithm

Figure 6: System view of real-time fuzzy scheduler

5 CONCLUSIONS AND FUTURE WORK

In this work we have proposed a solution for deployment of e-Commerce concept in the fast food restaurant chain that brings more convenience for both service provider and clients. The novelty of the work lays in possible implementation of client prioritization based on a well-known computer science concept of operating system task scheduling, that will work the best under the assumption that we have enough information considering stochastic component of service time. For such a credible statistics we need to have the insight of a system in a long run. The proposed system brings numerous benefits to both parties involved in the service process.

The system will boost the efficiency of the service by eliminating ordering waiting time and will cut costs by decreasing staff needed for order acceptance and delivery performing. It will also increase the visibility (if coupled with Internet - browsing and positioning) of the services, and enables better demand insight that brings more flexibility. Moreover, it will enable market profiling and targeted profiling, and gives the possibility to offer more e-services according to clients’ need.

There are many benefits for the clients also. They will achieve precise insight in offer and service condition, the queuing time will be eliminated and they will have the possibility to get delivery in short time as prioritized users. They will also have more payment options, and multilingual interface.

Our future work will focus on collecting and thoroughly analyzing statistical data. In the early phase of system deployment, realistic assumptions for worst case scenario needs to be made. Also, tuning of scheduling algorithm needs to be performed with a detailed testing with realistic data versus other scheduling schemes.

REFERENCES

MOBILE BUSINESS EXPERT ADVISOR

Danco Davcev, Marjan Arsic and Dalibor Ilievski
Faculty of Electrical Engineering and Information Technologies
University “Ss Cyril and Methodius”, Skopje, Macedonia
eftdav@feit.edu.mk, amarjan@on.net.mk, i_dalibor@yahoo.com

Keywords: Mobile Expert, on-demand knowledge, Web services, XML Agents, Fuzzy logic, Pocket PC.

Abstract: In this position paper, we introduce the service which enables consultations by the Experts and/or web services via mobile devices. The advanced development of wireless networks and mobile devices with various connection features made a great substrate for the development of services which are based on immediate response. One such service is a service for on-demand knowledge. A Helpdesk operator can post a request for consultation to the Experts and/or web services. The goal is to make knowledge available on demand at any time and any place. This approach gives faster problem solutions, more productive expert and/or web services, and high availability of the knowledge.

1 INTRODUCTION

The benefit of mobile devices combined with stable and cheap wireless networks made new kinds of services to appear. Experts are becoming more available than ever. Their services can be easy utilized and customers will have direct answer to the required question. In some cases (if appropriate) one of the XML agents communicates with the necessary Web services to find the solution of the problem.

Exchange of knowledge and consultation process among customers and available expert authority via helpdesk operators are very important aspects of quick problem solution using the mobile devices.

The goal of the mobile technology is to make information instantly available to customers. The new business processes can benefit a lot by using mobile connection between customers and experts. The exchange of the various content types of media data (text, drawing, and sound) will enhance the efficiency of consultations between the customer and the expert advisor. It gives benefits in reducing of costs, avoiding future problems or possible claims, saving time, increasing profit and customer satisfaction.

There are many desktop applications that support multimedia communication among participants. Communication among several participants that includes multimedia transfer is at the beginning of the development and utilization in the sphere of mobile handheld devices. Hence, new demands are imposed to these devices: greater processing power to support real time multimedia transfer (video stream, voice stream, and file exchange), greater memory space. Modern feature rich applications demand space on the screen for displaying all available features, which could be an issue when handheld devices are used.

A great consideration should be given to the readability and utilization of small mobile user interfaces. In this paper, we present our approach in creating a Mobile Business Expert Advisor (MBEA) for demanding and exchanging of various content types and knowledge on mobile devices. The interface of our MBEA adapts the dimensions and visibility of the user controls according to the user’s preferences and utilization of the controls and also, the current usage of the specific media contents within the business process. Our knowledge-based interface is managed by fuzzy logic and many XML based agents.

The related work in the second section gives some comments to similar MBEA-s. The architecture of our Mobile Business Expert Advisor is elaborated in the third section, which is the main contribution of this paper. The fourth section describes the design of MBEA. Implementation and Evaluation of MBEA are presented in fifth section. Finally, in the sixth section we conclude the paper.
2 RELATED WORK

The agent-based approach that uses fuzzy logic to determine importance of certain information is elaborated in (David Camacho, 2001). The agents described in this paper are based on access to the Web. We use similar approach to determine importance of user interface features in regards to user preferences and media contents used in current session. However, in our approach is used wireless network access for communication between handheld devices.

In (Vlado Glavinic, 2007), an intelligent tutoring system was presented. They use agents to recognize the device and the way user device connects to the global network and according to that information, a learning content is adjusted to the particular device and sent to the learner’s device. It differs from our system because we introduced agents whose purpose is to intelligently adapt the user interface to the preferences of each user.

Intelligent Mobile Answering Service is given in (Business Wire, 2006). Customers can use free-form questions on any subject with natural language queries. They can send text messages containing questions in any form just as easy as they send text message to their friends and receive a specific answer to their mobile phones. The natural language queries allow customers to ask questions in the way they normally express themselves. Unlike this, our approach is based on resolving business issues in various format types by the experts.

In (Lu, L. Kitagata, 2003) an agent based adaptive user interface control for desktop applications is presented. Although we share similar ideas, our approach is multi-agent based adaptive interface which uses fuzzy logic for handheld devices.

In (Weichang Du, 2006), collaborative applications can be built using two types of agent collaborations, agent communications through XACL and agent visiting. Agents are represented as XML entities, not programming language entities. Secondly, agent hosting services are implemented as web services with published WSDL, not programs in certain programming languages with published APIs. Thirdly, although XML agents’ behaviors have to be coded in some supported languages, the interactions between agents and hosts in agents’ behavior code are through invoking local hosts’ web services, which is neutral to programming languages and host operating systems.

Although we also use XML entities as agents and we share similar ideas of using web services (when appropriate), in our approach we use more flexible protocol for communication among agents which is based on fuzzy-based knowledge.

In (Pasquale De Meo, 2007), an XML-based multiagent recommender system for supporting online recruitment services is proposed. Although the main purpose of this system is to provide Online Recruitment Services, it is also agent and XML based and as a consequence, it can easily cooperate with company information systems. The so called ontology of their user agent stores the profile of a given user concerning the job search. It is different from our MBEA because we plan to build a general purpose advisor system (not only for recruitment services). For the time being, our objective is to have a self-adaptive system from the point of view of an efficient communication between user and the advisor according to the user’s preferences and the current possibilities of the communication system.

In (Zhiyong Weng, 2007) a feasible framework that combines agent mobility and intelligence for consumer-oriented e-business applications is proposed. This framework complements the current Web-based systems by adding the wireless channel of mobile agents. In our work the mobile agents use an adaptive communication protocol based on fuzzy logic. In addition, our framework includes web services.

3 MOBILE BUSINESS EXPERT ADVISOR (MBEA) ARCHITECTURE

There are three kinds of users like customers, helpdesk operator and experts. Helpdesk operator is the link between experts and customers.

Figure 1: Architecture of the MBEA.

The goal of the helpdesk operator is to response any kind of requests on demand by the customers in
appropriate formats (text, draw, VoIP) and to send the complex requests to the experts and/or web services which should be resolved by them. Client applications are implemented on Pocket PCs.

There is only one helpdesk operator and many experts in one consultation session.

The network communication is realized in a way that clients send UDP datagrams to the server, and the server resends the UDP datagrams, according to the contents of the received message.

The communication protocol between the agents and the information flow rely on fuzzy logic. In this case, the fuzzy logic is used for enabling or disabling (adaptation) some controls depending on the quality of the service (QoS), i.e. the signal strength. So, if the signal strength is good, the features of chat, draw and VoIP will be enabled. However, for medium signal the VoIP feature will be disabled. For bad signal, beside the VoIP, the draw feature should be also disabled, and the only active communication feature will be chat.

Streams are used for the voice transfer. Desktop based clients (for both helpdesk operator and experts) can be also connected to the server.

The Web services are used to find the solution of the problem using the wireless Internet access. If the Web services are unable to find the solution, the Experts are activated. Some types of Web services support and convert various types of media data.

The experts interface has controls for chat, draw, VoIP, file upload/download, authority control (VoIP channel).

The helpdesk operator can give the control over the interface (or by giving the speaking possibility) and reclaim the control later. Helpdesk operator has the role of a moderator in the consultation process.

Figure 1 shows the architecture of the MBEA. Client applications reside on pocket PCs and they have three main parts: Helpdesk agent, which manages operator interaction within the application and communicates with interoperability agent. VoIP client serves for the voice transfer, and the client application represents all the features mentioned above (chat, draw, voice, file up/download, communication among clients). Expert Agent receives and sends the resolved requests using various file types (like text, image, and voice). Collaboration Agent calculates the cost of the corresponding expert’s service. The communication is realized through wireless area networks connected to the Internet.

The server contains Interoperability Agent which redirects the complex request to the Experts and/or Web services and exchanges reformatted media types; Passive helpdesk Agent calculates the helpdesk interface; Media Agent monitors the media contents (text, image or voice); Experts and helpdesk preferences are stored in XML data storage; VoIP Server is responsible for voice transfer and UDP server application manages the login, chat, draw and control. For the purpose of saving the memory and processing power of the clients handheld devices, Passive Helpdesk Agents and Media Agent reside on the server side. The agents are XML based.

4 MBEA DESIGN

In this paper, we describe our approach in development of a MBEA by using multi XML agents and fuzzy logic. Helpdesk agent resides on the client side, and monitors the helpdesk operator’s interaction by using the features of the application and communicates with interoperability agent (e.g. sends the request to the available competent experts and/or web services, receives the response in corresponding format and regulates the customers billing). The triggers that demand helpdesk agent’s actions are: drawing, sending chat messages, and communicating using VoIP. This agent is named Helpdesk because it helps the customers to simply resolve their requests. At the start of the communication session, helpdesk agent communicates with the Passive helpdesk agent and interoperability agent that reside on the server. The Passive agent demands the information from the Media agent about the type of the files to be used for the particular communication session. This agent is named passive helpdesk since it waits to be invoked; it recalculates the interface structure and sends the data to the helpdesk agent which finally draws the interface. The Expert Agent receive the requests in some format (text, file, voice) from the interoperability agent, concerns about the solution of the requests, sends the solutions to interoperability agent, modifies the expert’s profile, checks the QoS and billing status.

The main tasks of the interoperability agent are: receiving the requests from helpdesk agent and forwarding them to the expert agents and/or web services. Also it exchanges reformatted data files with helpdesk agent. Media agent monitors the folder with the data files for the communication session, and sends this information to the Passive helpdesk agent. The Passive helpdesk agent recalculates the helpdesk interface according to the
information from the Media Agent, the history of interface affinities (chat_aff, draw_aff, sound_aff, QoS) for the particular user and the helpdesk agent’s message about the use of the features. Then, it sends the corresponding information (for the type of the helpdesk interface) to the Helpdesk agent which adapts the helpdesk interface according to this information. The collaboration Agent calculates the time session for realization of the expert task and the cost of the service.

Helpdesk and Expert profiles with personal information about the qualification and the knowledge for all of them are stored in the XML database. Additionally a history of affinities is stored for the helpdesk operator and the experts. According to the history, Passive helpdesk agent adapts the helpdesk interface on the beginning of the communication session.

There are currently three predefined user interfaces, which can be preloaded on the client’s device, according to his preferences and the media contents for the current communication session. The dimensions and the positions of the controls would be stored on the helpdesk’s affinity table on the server. One kind of an interface displays chat on the larger part of the screen, the second interface has larger drawing view on the screen, and the third one has equal space for the chat and the drawing space. VoIP feature doesn’t occupy much space on the screen, so the buttons for voice are always displayed and they will be active if the Signal Strength is satisfied, i.e. QoS is strong.

Since there is no universal conclusion for making decisions of which user interface should be preloaded, fuzzy logic approach is used to model that kind of the imprecise information.

The interaction made by the helpdesk operator using the features of the application (chat, draw and voice feature), is represented as a vector (chat_aff, draw_aff, sound_aff). A linguistic variable named RESULT_PREFERENCE is introduced, and it accepts values from the set of terms {increase_chat, increase_draw, the_same}. This variable represents the visualization of the helpdesk interface, produced by (1) the helpdesk operator interaction with the interface; (2) the contents of media folder and (3) the quality of service of the signal strength that decides more accurately which interface to be preloaded.

The class diagram for WiFi signal strength is shown on Figure 2. Two classes are developed, WiFiPeek and SignalStrength. The WiFiPeek class implements all the Wi-Fi query related elements. The class uses the NDIS User mode I/O driver (NDISUIO) to perform Access Point (AP) scanning.

The GetAdapters function can be used to query names of network adapters. It calls the built-in NDIS (not NDISUIO) driver. The function fills a buffer with adapter names separated by commas. The RefreshBSSIDs function requests that the driver initiate an AP survey. It takes one parameter: an adapter name. The GetBBSIDs function returns the list of available stations, i.e. peers and Access Points. The function getPreferredMACAddress returns the MAC address of the connected (associated) Access Point.

In order to create the MBEA for calculation of the user interface visual features, the following fuzzy variables are defined for this expert system: U_A (user affinity) which presents the most used feature by the user, i.e. it has the greatest affinity for the feature; M_P (media profile) presents the profile of the interface that should be used according to the media contents; QoS (Quality of Service) presents the strength of the wireless signal. Output variable named I (action for interface adaptation) presents the necessary interface to be preloaded onto the user’s device screen, according to the input variables.

The linguistic variable Quality of Service (QoS) accepts values from the set of terms {Strong, Medium, Weak}.

The linguistic variable U_A accepts values from the set of terms {Chat, Draw, VoIP} and it has normalized values of utilization between 0 and 1 (0% to 100%). It is equivalent to the number of times of usage of the features chat, draw or VoIP respectively.

The linguistic variable M_P accepts values from the set of terms {Text, Drawing, Sound}. It has normalized values of utilization between 0 and 1 (0% to 100%), which is equivalent to the number of text, drawing and sound files in the media profile respectively.

In the table 1 is presented the simulation for fuzzy linguistic values for U_A, M_P and QoS and...
the value of I is calculated, based on the fuzzy rules. The action needed to adapt the interface is described with the following XML code like the example 1 of the Table 1.

```xml
<IF U_A="DRAW" M_P="DRAWING" QoS="WEAK">
  <ASSIGN I="THE_SAME" />
</IF>
```

In the first example from the Table 1, the draw affinity is more utilized than chat and VoIP and the media profile has more drawing objects than text or sound. It is expected to be increased the drawing interface, but because the signal is weak and the drawing features are not supported on such signal, the interface stay unchanged (the same). The total numbers of fuzzy rules are 27, the combination of the values of all linguistic variables. The next examples are similar like the first one.

Table 1: Some examples for adapting interface using fuzzy linguistic variables.

<table>
<thead>
<tr>
<th>Example</th>
<th>1</th>
<th>2</th>
<th>3</th>
<th>4</th>
<th>5</th>
<th>6</th>
</tr>
</thead>
<tbody>
<tr>
<td>User Affinity (U_A) Utilization %</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CHAT</td>
<td>30.00%</td>
<td>57.00%</td>
<td>57.00%</td>
<td>11.90%</td>
<td>39.10%</td>
<td>41.90%</td>
</tr>
<tr>
<td>BROWSER</td>
<td>23.00%</td>
<td>23.00%</td>
<td>23.00%</td>
<td>28.20%</td>
<td>39.60%</td>
<td>23.20%</td>
</tr>
<tr>
<td>VOIP</td>
<td>15.00%</td>
<td>45.00%</td>
<td>45.00%</td>
<td>45.00%</td>
<td>36.40%</td>
<td>34.00%</td>
</tr>
<tr>
<td>Media Profile (M_P) Utilization %</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TEXT</td>
<td>26.07%</td>
<td>38.30%</td>
<td>38.30%</td>
<td>28.40%</td>
<td>29.80%</td>
<td>76.10%</td>
</tr>
<tr>
<td>DRAWING</td>
<td>50.00%</td>
<td>63.05%</td>
<td>63.05%</td>
<td>34.21%</td>
<td>30.91%</td>
<td>14.29%</td>
</tr>
<tr>
<td>AUDIO</td>
<td>23.73%</td>
<td>55.15%</td>
<td>55.15%</td>
<td>26.32%</td>
<td>30.89%</td>
<td>9.57%</td>
</tr>
<tr>
<td>QoS Signal Strength %</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>STRONG</td>
<td>0.00%</td>
<td>0.00%</td>
<td>0.00%</td>
<td>100.00%</td>
<td>100.00%</td>
<td></td>
</tr>
<tr>
<td>MEDIUM</td>
<td>10.00%</td>
<td>25.00%</td>
<td>30.00%</td>
<td>100.00%</td>
<td>8.00%</td>
<td>0.00%</td>
</tr>
<tr>
<td>WEAK</td>
<td>30.00%</td>
<td>75.00%</td>
<td>0.00%</td>
<td>0.00%</td>
<td>0.00%</td>
<td>0.00%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Action for Interface Adaptation (I)</th>
<th>Fuzzy Value %</th>
</tr>
</thead>
<tbody>
<tr>
<td>INCREASE CHAT</td>
<td>30.00%</td>
</tr>
<tr>
<td>INCREASE DRAW</td>
<td>10.06%</td>
</tr>
<tr>
<td>THE SAME</td>
<td>30.00%</td>
</tr>
</tbody>
</table>

The interaction among agents is shown on the Fig. 3.

5 MBEA IMPLEMENTATION AND EVALUATION

Applications for Wireless MBEA are developed in C++ Visual Studio .NET 2008 development framework with using of the MFC (Microsoft Foundation Classes) library. Operating systems used on the pocket PCs are Windows CE, Windows Mobile. The clients which reside on desktop PCs and the server use the standard .NET Framework. Operating system for desktop machines on which the application is practically deployed, is Windows XP SP2. The interface adaptation is realized on different screen resolutions.

The two instances of MBEA interface are shown on Figure 4 for two different screen sizes of Pocket PC-s.

The initial feedback of MBEA was given by 20 colleagues in ICT. They were asked to use the system for stock exchange analysis in last week. General opinion among participants for the interface...
usability, functionality and visibility is average (80%). Interface is functional and suggestions for interface rearrangement are mostly done according to the user needs.

The questions that examine interface usability, functionality, visibility and provide information for future upgrades are:
1) Are you satisfied with the way of adaptive presentation?
2) Can you easily select the parameters of MBEA?
3) Were the results of the MBEA solutions clearly displayed?
4) How much the MBEA helped in the business process?
5) Does the MBEA satisfy the quality of the service?
The questions are answered with “Yes”, “No” or “I Don’t Know”. The answers are given in the Table 2.

<table>
<thead>
<tr>
<th>Question</th>
<th>Q1</th>
<th>Q2</th>
<th>Q3</th>
<th>Q4</th>
<th>Q5</th>
<th>Average Percent</th>
</tr>
</thead>
<tbody>
<tr>
<td>Yes</td>
<td>17</td>
<td>15</td>
<td>16</td>
<td>14</td>
<td>18</td>
<td>80%</td>
</tr>
<tr>
<td>No</td>
<td>2</td>
<td>3</td>
<td>1</td>
<td>4</td>
<td>1</td>
<td>11%</td>
</tr>
<tr>
<td>I don’t know</td>
<td>1</td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>1</td>
<td>9%</td>
</tr>
</tbody>
</table>

6 CONCLUSIONS

This paper presents our approach that uses XML agents and fuzzy logic in order to achieve adaptive MBEA. We have developed such mobile system which determines the features of the interface according to the contents for the specific session. Fuzzy logic approach is used for the communication protocol between XML agents and for interface adaptation. It can be also used to dynamically reconfigure the interface according to the preferences and the type of request. In this way, we proposed an adaptable MBEA for handheld devices which brings benefits to the developers of applications for this software environment, to users of mobile business systems and at the end, to all mobile device users. General opinion among participants for the MBEA usability is positive. In the future work, we plan to provide an detailed analysis of the user’s feedback.

REFERENCES


A MOBILE BUSINESS PROCESS DEPLOYMENT FRAMEWORK FOR DEVICE INDEPENDENCE AND CONTEXT-AWARE ENVIRONMENTS

Torab Torabi, Saqib Ali and Hassan Ali
Department of Computer Science and Computer Engineering
La Trobe University, Melbourne 3086, Australia
t.torabi@latrobe.edu.au, ali.saqib@gmail.com, ali.hassan@gmail.com

Keywords: Business processes, Location aware, device independence, context-aware.

Abstract: Deployment of a business process to mobile devices in different context and in a device independence environment is a challenging task. The desirable is that a Business Process could be deployed to any mobile device in any location or environment the user operates. One obstacle in business process deployment on mobile devices is the rapid changes in mobile technologies. In this research we have developed and implemented a mobile business process deployment framework that caters for both device independence and context-aware environments. In this paper the focus is on how device independence and context-awareness can be integrated for mobile business processes deployment.

1 INTRODUCTION

The companies recently have been paying a great deal of attention to the potential of mobile communication technologies to redefine and extend the world of traditional E-Business by making its applications more available to the mobile users. According to (Swaminathan and S.R. Tayur 2003), E-business is defined as “a business process that uses internet or other electronic medium as a channel to complete business transactions”.

Mobile Business is often described as the successor of electronic business and defined as the subset of it (Adam, Chikova et al. 2005). It is assumed that any business operations performed by desktop computer can also be performed via wireless network. However mobile technology offers additional possibilities that are unique to the wireless world and cannot be performed via fixed network. For example, providing the mobile users with the services of location aware and context-aware applications which cannot be performed with a fixed internet connection (Zhang, Archer et al. 2003). Mobile business can be defined as the exchange of goods, services, and information using mobile devices (Paavilainen 2002).

Mobile Services are usually characterized by mobility; reach-ability, localization, and identification (Kaasinen 2003) (Siau and Shen 2003). Mobility is the central distinguishing feature of mobile technology. As mobile devices constantly accompany their users, people can receive and send data regardless of place and time. They can also be reached by people at all times. This feature is especially useful to logistic and supply chain companies that regularly need to reach their mobile workforce to allocate various tasks to them.

The success of the recent applications is based not only on new technology, but rather on its proper use (Liang and Wei 2004). Wireless applications would be advantageous only when they can be useful to their end users (Kalakota and Robinson 2001). Mobile business value proposition originate from the fusion of the wireless technology with already available electronic business applications. On the other hand, the unique features of mobile business – mobility, localization give rise to emergence of completely new applications and business models.

1.1 Mobile Business Processes

according to Davenport (Davenport 1992) a business process can be defined as “a specific ordering of work activities across time and place, with a beginning, an end, and clearly identified...
inputs and outputs: a structure for action.” The business profit depends on efficient delivery of goods and services controlled by business process (Ali, Torabi et al. 2006) (Ali, Soh et al. 2005). So there is a need for the companies to make use of the technologies to make their product more profitable and their services more efficient.

Many companies have been able to make their traditional business processes into mobile business processes. “Mobile Business Process” is a business process, when a place of execution of an activity can be different in different instances of the business process or places can change during the execution of an activity (Ritz and Stender 2003) (Berger, Bouzid et al. 2003). Mobile Business Processes can be based on these three assumptions:

- “Uncertainty of Location”.
- “Uncertainty of Location” is externally determined, and
- A corporation with external resources is needed in the execution of the process.

1.2 Context-Aware Mobile Applications

Context means situational information, or as (Barkhuss and Dey 2003) states: “Context is any information that can be used to characterize the situation of an entity. An entity is a person, place, or object that is considered relevant to the interaction between a user and an application, including the user and the application themselves.”

Context-aware is a concept where the applications can discover and take advantage of contextual information such as user location, time of day, nearby people, devices, and user activity.

An example of context-aware applications is the scenario in which a consumer makes use of the mobile technologies in retail grocery supply chain through mobile shopping of electronically referenced grocery products. This kind of application is also beneficial to retailers, who by knowing the exact location of the consumer can allocate the workforce more accordingly and efficiently, and can avoid out-of-stocks (Hakkila and Mantyjarvi 2005) (Koolwaaij and Strating 2003).

1.3 Device Independence

In today’s world of rapidly changing technologies and the rise in the use of mobile devices has created a need for content adaptation. According to one of the authors (Mikhalenki 2004), the goal of device independence is to develop ways for future web content and applications to be authored, generated, or adapted for a better user experience when delivered via many device types.

Currently many companies typically design web applications for desktop browsers. And adaptation of this kind of web site for small display is effectively impossible and the companies creating a parallel site for these devices is also impractical. To solve this problem, the device-independent approach is used to support different devices without the high cost.

Device Independence technologies can be divided into three different categories namely: -
- intermediate, client-side and server-side (Butler, Giannetti et al. 2002).

Intermediate approach can offer limited adaptation to the content delivery. This approach gives data-enabled phones access to web sites either omitting server’s full resolution colour images or changing it to the low resolution depending on the device display capabilities (Hwang, Kim et al. 2003) (Butler, Giannetti et al. 2002).

In the client-side approach the content adaptation can occur on the device itself. The advantage for this approach is that the adaptation code has direct access to the device capabilities (Butler, Giannetti et al. 2002).

Server-side content adaptation approach offers maximum control over the delivered content including the ability to change content, navigation and style. In this approach the server is assumed to have sufficient information about the delivery context, including the delivery device capabilities.

Using device independence approach the content integrators do not need to develop content for every single device, where the content can be delivered to different devices without accumulating more resources and effort.

In the next sub-section we will discuss different deployment methodologies and we will discuss our approach in mobile business process deployment.

1.4 Deployment

In terms of technology, the word “deployment” means “Installing, setting up, testing, and execution” (Ali, Torabi et al. 2006). Therefore, deployment can be interpreted as a general process that must be customized according to specific requirements or characteristics.

Deployment of software, applications or a process is a complex task which covers all the
activities from the end of the development itself to
installation and maintenance of the application on
the consumer devices. In (Richman 2001) the
authors have compared different types of
deployment techniques in terms of their scale,
complexity, expressiveness, and barriers to first use.
The deployment solutions being handled by these
implementation techniques: manual, script,
language, and model-based deployment.
The automation of application or service
deployment improves correctness, speed and
documentation but, as different companies have
experienced, it comes at an increased cost in
development time and a steeper administrators’
learning curve.

According to recent research, business processes
are deployed through XML web services. Web
services can be considered as the emerging
distributed middleware technology that uses a
simple XML-based protocol to allow applications
to exchange data across the web (Zhang, Zhang et al.
2004; Hammadi, Ali et al. 2006). At the core of the
Web Service is the Simple Object Access Protocol
(SOAP) an XML-based communication protocol for
interacting with Web Services. The SOAP
specification includes syntax to define messages,
encode or serialize rules for data exchange and
conventions for representing RPCs (Austaller,
Kangsharju et al. 2004).

WSDL is used for describing the services
available. It describes where the service is located,
what operations are supported and the format of the
messages to be exchanged based on how the service
is invoked (Gokhale, Kumar et al. 2002; Austaller,
Kangsharju et al. 2004). On the other hand Business
Process Execution Language (BPEL) is the language
that is used to implement business processes in Web
Services. It defines a notation for specifying
business process behavior based on Web Services
(IBM, Systems et al. 2002).

Certain technologies and models have been
presented in recent years for the deployment of
services and applications using mobile technology.
One of the technologies being used recently by the
companies for the deployment of services is over-
the-air (OTA) deployment. OTA is becoming
increasingly important to support. OTA delivery
enables easy deployment and upgrades to the
applications, thereby reducing the disrupting effect
which installation of new applications and upgrades
may have on mobile users (Taconet, Putrycz et al.
2003) (Fjellheim 2006).

Another kind of deployment technique is “static
deployment” where the user connects to the site of
the application server through its mobile and then
subscribes to a download operation. Then the user
receives an SMS containing instructions for
downloading and installing the application.

The framework known as Smart Deployment
Infrastructure (SDI) is designed to facilitate the
installation of large distributed applications for any
kind of user terminal. This framework is also
presented for the context-aware deployment of
applications to the mobile users (Taconet, Putrycz et
al. 2003). The framework is implemented with
middleware technologies like CORBA and SOAP
which facilitates the development of large scale
distributed applications. SDI offers automatic
deployment of multi-component applications and
provides a deployment solution to customize
installation and to adapt to device capabilities
(Courtney 2000; Taconet, Putrycz et al. 2003).

In the next section we will present a framework for
mobile business process deployment. The
framework caters for deployment of a process in a
device independence and context-aware
environment.

2 THE FRAMEWORK

In this section we present a novel framework for
business process deployment in a mobile
environment taking into account the context of the
user and specification of the device to be used. The
overview of the architecture is shown in Figure 1.

The framework proposed consists of five
different components:

• Deployment component
• Process specification component
• Device Specification component
• Context-Aware component
• Mobile Interface component

The “deployment component” is the main
component of the framework. This framework
consists of two layers namely Composition Layer
and Deployment Layer. The work for the
composition layer is to compose a selected Business
Process, setting the context of the deployment and
customizing the process according to the device
being used in the “Mobile Environment”. Where the
deployment layer activates the customized business
process to the Mobile Environment.
The second component in this framework is “process specification component”. This is a traditional Business Process component consisting of activities, resources and user interfaces. User Interfaces are stored as XML documents in the UI Database associated with process specification component.

“Device Specification component” is a typical component. It consists of different hardware configurations, screen resolution settings and other features for the mobile devices being currently used in the Mobile Environment.

“Context-Aware component” consists of many parameters but in our Context-Aware component, we only consider two parameters “Location and Environment”. The location of the Mobile User is determined using different technologies like GPS, GSM, or MPS as shown in location server defined in our architecture. The “Location Database” is used to store all the locations of the users and later can be used in process deployment.

“Mobile Interface component” consists of the N number of users using N number of mobile devices. The Mobile devices being used can be pocket pc’s, smart phone’s or PDA’s etc.

In the rest of this section we will present the Framework Design and the System Architecture. The three-tier system architecture best shows the interaction between the components and communication between the three tiers.

2.1 The Framework Design

This section will focus on the Object-Oriented Design for the proposed Mobile Business Process Deployment Framework.

We will start our design with the "High Level Component Diagram" as presented in Figure 2.

Our Object Oriented design is generic, flexible, scalable, reusable, and it provides the flexibility to coordinate, coordinate, and synchronize between components and process automation.

This component diagram shows the system functionality through a high level of abstraction. For our design there are nine active components.

In this section we will be present a brief description of each of the components. We will briefly discuss the functionalities and properties that each of the components possesses then in the next section we will discuss each component in detail with the help of a class diagram. The brief explanation of each of the components is as follows:

Mapping Engine: - This component is to take care of mapping between different protocols. All the functionalities like displaying simple map, Geo-coding, Reverse Geo-coding and routing would be done by this component.

Process Specification: - This is a traditional Business Process component consisting of activities, resources and related user interfaces. All the specifications would take place in this component. The specification for the Business Process along with its activities, resources and user interfaces are stored in process database.
Device Specification: It consists of different hardware configurations, screen resolution settings and other features for the mobile devices being currently used in the Mobile Environment. In the framework these would be stored as an XML documents and would be used at the time of deployment. Device database is used to store all the device specification for different mobile devices.

Context-Aware: Consists of many parameters including "Location and Environment" parameters. In this component we show how the location and environment parameter can be determined using different positioning and sensor technologies. Context database is using to store the context information of the user in the mobile environment.

Deployment: This is the major component of the framework. This component is where the actual deployment would take place. All the components would be integrated and coordinated before deploying to the mobile environment. All the Protocols and Web Services technologies would also reside in this component.

XML Parsing Component: All the XML parsing of the incoming responses, XML documents stored for Process Specification, Context and Device Specification is carried out by this component.

2.1.1 Mapping Engine

The purpose of the mapping engine is to provide the framework with the maps if required during deployment. The maps are really important if one wants to show the user its location that in our case one of the parameters of the context information. The class diagram for the component is shown in Figure 3.

Mapping Engine is adopted from a Location Based Services Company, iGoPlus, this exercise. As can be seen from the class diagram above, there is a SDK Interface that links the application with the mapping engine.

Through this engine our framework would be able to do the following functionalities:

- Display simple map with centerOnAddress or CenterOnPosition. Where the map would be centered either by latitude longitude position or by address.
- Show Geocoded address on the Map.
- Do Reverse Geocoding and show the position of the points on the map.
- Display the route between two points on the map. This also includes displaying the Text Instructions required in navigating from start to end position.
- Do the map customization in terms of orientation, color, mapView and size.

2.1.2 Process Specification

This component takes care of the Business Process Specification along with its activities and resources. A Business Process can have many activities and each can have different resources.

User interfaces are attached to each subsequent activity. Not every activity may have a user interface. User interfaces are associated to activities during business process specification. The class diagram for Process Specification is shown in Figure 4.

A Business Process may have number of activities and each activity has some resources at its disposal. Among others other characteristics of resources includes whether or not they are
consumable or sharable resources. User interface specification is an important part of this component. In our framework we are assuming that user interfaces would not be specified but only referenced. Its attributes include id and path where user interfaces are stored. It should be noted that not all activities would have user interfaces attached to it.

The Process Specification is stored in the database. It can be seen from the diagram that there is one main super class "Database Management" which has all the subclasses from it, which are used in the system. The class "Database Management" contains all the attributes and the operations common to all the classes.

2.1.3 Device Specification

Having discussed the design for Mapping Engine and Process Specification, in this section we will present the design for Device Specification. Device Specification is stored as an XML document in the database. Device Specification consists of software and hardware specification. Other components interact with this component through Device Interface. The design of the device specification of the framework is shown in Figure 5.

2.1.4 Context-Aware

In this section we present the design for context-aware component. There are many parameters as part of context-aware information, including the location and environment. The contextual information for this component would be stored as XML documents in the database.

The design for this component is based on the working of commercial applications like on-board navigation systems, gadgets or sensors for light or noise. The context parameters, location and environment will be determined as shown in Figure 6.

2.1.5 Deployment

Having specified Business Process, activities and resources, the deployment component brings context information and customized user interfaces together before deploying it to mobile environment. The composition layer coordinates and integrates all three components and the mapping engine component before the protocol layer can deploy it. The design for the deployment component is shown in Figure 7.
This component is the backbone for the deployment process. The Composition layer coordinates and integrates all the three components together towards the deployment process. Examples of functions being performed by this layer are as follows:

- `getLocation()`: Gets the location of the particular user from the Context-Aware Component. The response returned would be in terms of latitude and longitude.

A Business Process can be deployed either by HTTP connection or by SOAP commands. The Framework caters for both of these protocols.

### 2.2 System Architecture

The system architecture for the Mobile Business Process Deployment Framework is divided into three main parts client layer, business logic and database layer. The System Architecture is shown in Figure 8.

The client layer or tier interfaces between the system and the user. It consists of a mobile device running on operating system windows CE 5.0. The Business Process is deployed to a mobile device in form of XML document. The communication between client tier and business logic tier is done through XML web services hosted on ASP.Net Web Server.

Apache Cocoon Server is used in the framework to determine the software capabilities of a mobile device from the responses received from the client tier. ASP.Net local Web Server is used to support XML Web Services and Simple Object Access Protocol (SOAP). These technologies set up the communication between the server and the client.

Contextual Information and specifications for mobile devices are also stored as XML documents in the Mysql Database.

Deployment Engine Component consists of two layers Composition Layer and Protocol Layer. The responsibility for the composition layer is to compose a selected Business Process, setting the context of the deployment and customizing the business process according to the device being used. The protocol layer would set protocols for the deployment like setting up parameter for SOAP commands and setting up HTTP connections.

The database layer consists of database and Mapping Engine. For communication between Business Logic Tier and Mysql, there is MyODBC Interface that takes the request from the server and sends it to database.

### 3 CONCLUSIONS

In this paper we have discussed deployment of a business process to a mobile device in a chosen context and in a device independence environment. A Business Process can be deployed to any mobile device in any location or environment the user may be. This framework helps companies and developers to deploy same business processes to different mobile devices without any software or user interface configuration. Much research has been done in the field of context-aware and device
independence, our contribution to this research is to bring these aspects together and design a framework that can bring more flexibility to the provider and the consumer.

REFERENCES


GIS-BASED MAP GENERATION USING NEW SURVEY TECHNIQUES

Balqies Sadoun and Omar Al-Bayari
Surveying and Geomatics Engineering Department, Al-Balqa' Applied University, Al-Salt, Jordan
balqiessadoun@yahoo.com, obayari@bau.edu.jo

Keywords: Geographic information System (GIS), Digital Elevation Model (DEM), Ground Control Point (GCP), Three Dimensional (3D), Global Positioning System (GPS), Real Time Kinematics (RTK), Orthophotography.

Abstract: Conservation of historical sites depends on acquiring all data related to their life history and their shape. We are using modern survey technologies such as GPS, satellite images, photogrammetry and classical instrumentation to build a Geographic Information System (GIS), digital library and precise base map with all its embedded benefits at a low cost. GIS offers digital documentation of the site and its surrounding environment, which is essential for the conservation and protection process of any heritage. In this work we are presenting GIS, digital base map and an orthophoto for a newly established museum and its surrounding to be used in GIS modeling. Global Positioning System (GPS) precise measurements and a high resolution satellite image were used to produce a Digital Terrain Model (DTM) and an orthophoto for the site in order to create a three Dimensional model (3D). Finally, GIS and (3D) texture model for the museum were produced to conserve, protect, monitor and manage the facility. It offered as well, the scheme to create a web site to post all acquired information to publicize the new museum and the ability to employ any modern location based service (LBS) technologies (emergency, security, guiding etc.).

1 INTRODUCTION

It is very important to use modern technologies such as satellite images, remote sensing, photogrammetry and GIS (Grosman, 2000) to conserve cultural and historical sites. New survey technologies are very valuable in the mapping and the management of the archeological data (Kvamme, 1999), (Bewley and Raczkowski, 2002).

The objective of this work is the development of a base-map: to display a detailed map of the site and its general components, in order to portray an in-depth data of any selected part and to create 3-D model with the possibility of a walk through visit so as to conserve and manage the site efficiently.

To achieve our objectives we had produced a Digital Elevation Model (DEM) and orthophoto using some satellite images and a photogrammetry software. Then, the DTM produced is used to create 3D model (Baltsavias et al., 2001). Using digital photogrammetry and computing capabilities we were able to digitally document and 3D model the cultural site. The produced digital map by means of satellite images, digital photogrammetry and GIS, will offer endless of working possibilities (Bayari, 2005).

The site we intended to conserve is a museum located to the west of Amman, Jordan which was the house of a national hero. It is beauty and significance is due to its location and the Islamic arcade architecture of its quarters. It is the first of its kind in the country and need to be digitally conserved to allow the usage of modern LBS technologies in its operation. Satellite images were used to extract features with real dimensions and a base-map was extracted from a topographic map and orthorectified Image for the site. Moreover, a three dimensional texture model for the museum was built using field survey and GIS starting from topographic planner. Real measurements were translated to a database in a GIS environment to build the 3D solid model. Then, sketch up texture mapping software was used for building texture based on capturing field photos. Using Arc GIS environment we provided a real 3D model with true virtual reality model to allow all 3D GIS spatial analysis. We
designed a website that offers all the needed information (Location, base-map, photos, videos of the 3D model etc.). GIS has a great Capability to integrate and update graphic and non-graphic data according to user inputs to produce all kind of maps.

2 METHODOLOGY AND SOFTWARE USED

A Geographic Information System for the site was built to house all the digital data to allow it's updating, posting on the web and the usage of modern LBS technologies. The digital data allows better analysis, presentations, updating in addition to the production of 3D GIS texture model and a web site.

Several Software packages were used in the process including:

- **Auto Desk Land Desk Top 2004**: is used to create DTM, feature layer and data base, in order to produce a topographic plan for the site and to export AutoCAD layers to shape files to use in Arc Map.
- **ENVI 4.3** (the Environment for Visualizing Images) is used to create a high-resolution color image; PC sharpening fuses high-resolution panchromatic imagery with multispectral imagery and to correct the images from the distortion by Orthorectification algorithm.
- **ArcGIS (version 9.1)**, is used to Produce Base-Map from digitizing the ortho and classic map. 3D model was built in ArcScene depending on the attributes tables (extrusion the building according to height). Finally it Generates Triangulation Irregular Networks (TIN) for the site and compose all the data (images, 3D features, TIN, Sketch Up file, ENVI) together.
- **Sketch up Pro 5**: is used in exporting and importing 3D building from and to ArcGIS. It is a powerful tool for creating, viewing, and modifying 3D ideas quickly and easily. Although it had been developed for the conceptual stages of design, Sketch UP is a very powerful 3D tool for creating presentations.
- **Photoshop**: is used to eliminate any appearing effect on photos (shadow on facades, trees, humans and cars, etc.).

By providing measurements from the field and processing them we produced the topographic plane. Then, we used digital photogrammetry technique to orthorectify Quick Bird image to construct orthophoto. We used the topographic map and the orthophoto to produce a base-map. A solid 3D model of the building was produced in GIS. The sketch up texture mapping software was used to build texture based on capturing field photos, then converting the 3D building texture to Arc scene so as to produce 3D GIS Texture Model. Finally, the beautiful arcade house with the Umayyad Arabesque style (Fig.1) was turned into a digital data base museum.

A Topographic Survey of selected natural and built up features was done to produce the needed maps and plans with true features to scale. GPS Real-Time Kinematic (RTK) and total station provide identical output coordinate data to perform topographic surveys of the terrain, facilities, and infrastructure.

![Figure 1: Main Building.](image1)

![Figure 2: Topographic Plane for the museum site scale 1: 500.](image2)
features, stakeout methods, COGO functions, etc. COGO is a suite of programs aimed at coordinate geometry problems in Civil and Geomatics Engineering.

Land Desk Top 2004, field observations (1500 points) were loaded to produce the final topographic plane. Processing starts by adding the data, building the surface and the TIN. Finally, a topographic plan was created using Land Desk Top 2004 as shown in Figure 2.

Digital photogrammetry technique is used to Orthorectify Quick Bird image which was available as black and white multispectral, color, or PC-sharpened product. We chose 8 GPS points that can be used as GCP’s, and 5 points by Total Station from topographic plane. In order to orthorectify the image in GeoTIFF format, and files containing the RPCs for the Quick Bird Image, Digital Elevation Model (DEM) was used (Certer, 1988). All the data had been used as input to the ENVI software to produce the orthophoto high resolution 0.6m and multispectral RGB (Fig.3) and 3D surface (Fig.4).

3 BASE MAP GENERATION

A base map is the graphic representation at a specified scale of selected fundamental map information that is used as a framework upon which additional data of a specialized nature may be compiled. It is as well, a map from which other maps are prepared by the addition of information. To produce the base-map we had to do a topographic survey, and then to use a Quick Bird Orthorectified image to digitize roads network, agricultural features, buildings and finally to use a topographic map of Swieleh; a suburb in Amman. The collected data produced an orthorectified image and topographic plane. The flow chart of map generation is presented in Figure 5.

Map projection is a systematic transformation that allows the orderly representation of the earth's spherical graticule on flat map. Registration is based on a family of mathematical tools that are used to modify the spatial arrangement of objects in a dataset into some other spatial arrangement. The purpose is to modify these geometric relationships without substantively changing the contents of the data itself. The registration process involves changing one of the views of the surface spatial relationships to agree with the other, without concern about any particular geodetic referencing system.

Digitizers are the most common device for extracting spatial information from maps, photograph or other documents. In the Swieleh topographic map we digitized the contour lines and names of places. In the Orthorectified image we digitized the Agricultural lands (Fig. 6), Buildings (Fig.7), drainage pattern and roads network (Fig. 8).
As for the final base map layout (Fig. 9), we collected all Layers and added north arrow, legend, scale, etc...

Architectural Heritage Conservation is better achieved through building 3D photo-texture model integrated with GIS. The modernization of our world needs technical document production. Digital documentation of cultural heritage is not an end in itself but serves as a tool to make accessible information and better presentations.
In the process of 3D modeling and building the height database we used the digital maps to implement the field height data through GIS (Fig. 10).

Then, we applied the heights to their corresponding features to produce 3D solid model (Fig. 11) in GIS environment.

To build real texture 3D model under GIS environment (to conserve the real texture of all the facades), digital photos for the sides of the buildings were captured using a high resolution digital camera. Then, Photoshop software was used to eliminate any extra features captured in the photos such as trees, cars and people as it is presented in Figure 12.

Finally, Sketch Up software was used to export and import to and from ArcGIS (Using the SketchUp ArcGIS Plugin). The SketchUp ArcGIS Plugin was installed in the GIS environment to enable the transformation of 2D GIS data to sketchUp, seamlessly and transfer 3D texture model to an ArcGIS geodatabase. Figure 13 presents the real and the 3D texture model for comparison purposes.

### 4 CONCLUSIONS

GIS is an effective tool in managing, sorting, analyzing and presenting any architectural or other attribute data to serve in the field of conservation, management and the employment of modern technologies (LBS for emergency, guided tours, planning etc.).

- The production of a digital base-map by means of satellite images and digital photogrammetry is a new and fast technique, which allows the updating and the production of endless needed maps.
- The production of 3-D texture model helps in performing advanced analysis and studies. Processing topographic surveys and high-resolution satellite images is an excellent technology to conserve important sites accurately with real features.
- Finally, the creation of an important web site rich in information to allow global visitors, researchers and concerned groups is the optimal benefit of the work.

### REFERENCES


POSTERS
Keywords: e-Commerce, WebComfort, content management systems (CMS), e-commerce extensions for CMS.

Abstract: Electronic commerce (e-Commerce) involves a complex set of business workflows regarding the buying and selling of products or services by means of electronic systems such as the Internet. e-Commerce applications can be supported, amongst other alternatives, by content management systems (CMS) through specific extensions. This paper presents the eCT extension, an e-Commerce extension developed on the top of the WebComfort CMS platform, whose main goals are: (1) adaptability to different contexts, such as different types of stores, countries, cultures or tax systems; (2) extensibility, due to the ease of development and addition of new functionalities; and (3) ease of installation and configuration, to fulfill different business requirements. This paper discusses the major technical details of the eCT, such as: (1) the product facet model; (2) the support to different e-commerce models; and (3) the support for different payment, shipping and taxing methods.

# 1 INTRODUCTION

Electronic commerce (e-Commerce) involves business processes related to the buying and selling of products over electronic systems, such as the Internet. Electronic commerce is expected to continue growing at a rapid pace, and organizations are using it to reach a wider audience and create new business opportunities, through “electronic shops”, that offer products from a single firm, or from multiple individual stores (electronic shopping malls (Khosrow-Pour, 2006).

On the other hand, in the last years, a number of web-oriented content management systems (CMS) (Suh et al., 2003) has appeared, aiming to facilitate the management and publication of digital contents. These systems allow the collaboration of the various users involved in the process of creating and organizing digital content, and typically provide extension mechanisms that support the addition of new types of content as well as new languages and visual themes. These systems are providing organizations a way to support their needs for constant evolution and complex business requirements.

This paper is structured in five sections. Section 1 introduces the broad context of e-Commerce and CMS concepts. Section 2 introduces the eCT (the B2C e-commerce Toolkit for the WebComfort platform). Section 3 overviews of the eCT architecture and its extensibility features. Section 4 discusses some of the architectural decisions taken during the development of this toolkit. Finally, Section 5 presents the conclusions for this project so far, as well as the future work.

# 2 THE ECT TOOLKIT

The “e-Commerce Toolkit” (eCT for short) is an e-commerce extension for the WebComfort platform (WebComfort.org, 2008). WebComfort is an Enterprise CMS, promoted by SIQuant (SIQuant, 2008), developed on top of Microsoft ASP.NET technology.

In the WebComfort platform, a Module provides mechanisms to manage and layout a certain kind of content (e.g., images, text, links), and a Toolkit is an integrated collection of modules that provide specific functionalities. For example, the WebComfort Standard Toolkit defines a set of modules allowing the management and visualization of typical Web contents (e.g., images, text).
The eCT is a Toolkit for e-commerce support, more specifically for Business-to-Consumer (B2C) business model. From a quick perspective, the eCT allows the configuration of one or more electronic stores, that allow customers to browse the product catalog, add products to the shopping cart and proceed to the check-out. On the other hand, store administrators can configure and manage a number of store details (such as product catalog, inventory, promotions and orders) whereas eCT administrators manage aspects that are available to all the stores (such as languages and localization, currency units, shipping/payment and taxing methods).

3 ECT EXTENSIBILITY FEATURES/AREAS

The eCT is an extension to the WebComfort platform, aiming to support electronic stores that can be based anywhere on the world and available to a global audience. Due to the variability of possible contexts, the eCT must provide solutions for the different aspects that can vary (e.g., languages, payment systems, shipping systems, taxation, currencies, types of products to sell). So, the main principles that guide the design of the eCT are:

1. **Adaptability** to different contexts (such as different countries and cultures, types of stores);
2. **Extensibility**, namely by the ease of development and integration of new features;
3. **Ease of installation and configuration**, allowing a quick setup of eCT according to stores’ different business requirements.

This section discusses how these principles were taken into consideration during the eCT design.

3.1 Overview

The eCT aggregates a set of WebComfort modules that implement B2C e-commerce functionalities. The eCT has a multi-tier (three-tier) architecture, with the typical presentation, logic and data layers, with a WebComfort module being transversal to this tiers (see Figure 2). The presentation tier of a module contains the UI elements (Web User Controls and Web Forms) that present information to the user and interpret/respond to their actions.

The logic tier contains the module’s business logic. This level can interact with the WebComfort API, that provides a set of infrastructural features available to the WebComfort modules, like authentication or multi-language support. In a similar way, infrastructural features of the eCT are provided by the eCT API, that cover areas like: eCT configuration, store configuration, store operation management, catalog, inventory and promotions. The eCT configuration involves the configuration of global options shared by all stores; the store configuration involves the configuration of each store options (e.g., accepted shipping/payment/taxing methods, currency units); the store operation management involves the management of orders, customers and statistics; the catalog refers to the storing and management of product data; the inventory corresponds to the information of the physical storing of products; and promotions corresponds to the features that allow the promotion of certain products. At the eCT level, there are a number of possible extensions, namely: Product facets; Shipping; Payment; Taxing; Logging; Statistics and Reporting; and Currency Units.

The Data tier isolates the persistence aspects of the module. The eCT toolkit uses two data repositories, for the eCT data and statistics/logging data.

Now we present in more detail some of the more relevant aspects/features of the presented architecture.

3.2 Product Facet Model

One main requirement identified was the support for different types of products (e.g., books, digital files), which may differ in the way they are presented to the end-user, as well as in internal behavior. The eCT supports the notion of a “base product”, which can be specialized or extended by applying facets according to the context. These facets can be perceived as extensions to a product, that can add a certain behavior and/or data. For instance, in a fine arts portal we can allow the sale of fine arts originals by defining a facet “Fine arts original”, that allows the association of this information (author, materials, dimensions, etc) to the...
product. This model allows the association of various facets to a product, allowing the composition of facets to enrich it.

In addition of the physical products (books, clothes, etc.) supported by these extensions, the eCT supports also the providing of services, that differ from physical products in the sense that providing a service does not result in ownership, but results in some benefit to the customer. Examples of this can be found at tourism, entertainment, or cultural services.

### 3.3 e-Commerce Models

An important aspect of the eCT is its support for different business models, so that it can be applied to different domains and business contexts, namely (1) Store; (2) Shopping Mall; and (3) Shopping Mall with integrated shopping cart. According the "Store" model, the WebComfort portal should support only one store; according the "Shopping mall" model, the portal should support a set of independent stores, that manage their own specificities (catalog, orders and payments management); and according the "Shopping Mall with integrated shopping cart" model, stores are not totally independent, and should share the orders and payment management features (see Figure 3).

### 3.4 Payment, Shipping and Taxing Methods

The adaptation to different contexts mainly concerns the support of different payment, shipping and taxing methods. This is essential, because it is important to supply methods that fulfill the needs of different stores. To achieve this, it is important to allow the definition of new methods and the easy integration of these methods into the eCT. (Due to space restrictions other issues are not discussed here and the reader can consult (Baptista, 2008))

### 4 DISCUSSION

The majority of the aspects presented above are responsible for allowing the eCT to be a flexible extension for the WebComfort platform. Even maintaining the WebComfort approach of trying to keep as few dependencies between components as possible, namely dependencies between the eCT and the WebComfort platform and between the different components of the extension.

To support the identified requirements – adaptability to different contexts; extensibility; and ease of installation and configuration – a set of inter-related features were defined: the product facet model; the support to different e-commerce models; and the adaptation to different contexts in terms of payment, shipping and taxing methods.

To implement the product facet model we adopted the Decorator design pattern, that allows to add respon-
sibilities to individual objects dynamically and transparently (Gamma et al., 1995). This way, instead of trying to support from scratch all the expected product features in a complex class, it is now possible to define a relatively simple class and add features incrementally with decorators, being the final product the sum of all these facets. Consequently, this pattern allows the definition of a vast number of extensions (facets), that materialize specific product semantics. Regarding the behavior that allows the adaptation to different contexts in terms of payment, shipping and taxing methods options we opted for the Provider Model (Rob Howard, 2004), that is a mix of the Abstract Factory, Strategy and Singleton patterns (Gamma et al., 1995). In the eCT, this allowed the definition of an API that exposes the operations relative to the payment, shipping and taxing areas (as illustrated in Figure 4). For instance, the Payment API can have a method responsible for processing a payment, that contains no business logic; instead it simply forwards this call to the configured provider. The providers implement different payment methods, containing whatever business logic they require. Thus, custom providers for each of these services can be easily and independently developed and configured.

Figure 4: Overview of the provider model for payment, shipping and taxing.

5 CONCLUSIONS

Nowadays organizations and enterprises are using electronic channels to manage and publish digital contents as well as sell their products or services. In the context of our research, the e-Commerce Toolkit (eCT) was implemented, to support different kinds of electronic shops, either offering products from a single entity or from multiple individual stores (electronic shopping mall).

This paper presented and discussed some of the main features of the eCT, namely: the product facet model; the supported B2C e-commerce models; and the adaptation to different contexts in terms of payment, shipping and taxing methods.

For future work we plan to improve this toolkit, mainly in the areas of promotions and inventory, that do not yet have all the expected features, and to develop more Product Facets and more providers so that we can support a greater number of different contexts.

REFERENCES


SURVEY OF CONSUMERS’ DECISION MAKING PROCESS FOR ONLINE MUSIC SERVICE

Ayako Hiramatsu, Takahiro Yamasaki and Kazuo Nose
Osaka Sangyo University, 3-1-1 Nakagaito, Daito, Osaka 574-8530, Japan
ayako@ise.osaka-sandai.ac.jp, takahiro@eic.osaka-sandai.ac.jp, nose@ise.osaka-sandai.ac.jp

Abstract: This paper describes a survey of consumer behavior for online shopping, especially online music services, because they are the most popular online shopping service in Japan. Based on Howard's consumer decision model, questionnaires about decision making for online music services were given to 282 students. The questionnaire results show that almost 90% of these students have downloaded from online music services, and high school students use such services slightly more than university students.

1 INTRODUCTION

Internet penetration in Japan is remarkable. In 2006 its penetration rate was 68.5%, and the number of Internet users was estimated at 87.54 million (White paper, 2007). With the Internet’s diffusion, the popularity of e-shopping continues to increase. 63.7% of all Internet users have been using e-shopping for more than two years (Internet Association Japan, 2005, 2006), and these numbers are still growing. For marketing, understanding the behavior of e-shopping users is crucial.

To understand consumer behavior, user models are often constructed by analyzing the behavior data of users. Concerning Internet users, there have been some researches about online game models (Hsu and Lu, 2004) and e-shopping users (Shih, 2004) based on the Technology Acceptance Model (TAM) (Liu, 2007). Farag et al. also researched a model that analyzed relationships between e-shops and real shops (Farag et al, 2007). Moreover, Watanabe and Iwasaki analyzed why consumers buy PCs through the Internet (Watanabe and Iwasaki, 2007). However, decision making processes based on various factors and the relationship between them have not been analyzed much, due to such difficulties as direct observation and analysis of subjective condition of users.

Our research aims to understand how consumers make e-shopping decisions. Consumer behaviors for real shops (Evans et al. 2006) were researched. Based on Howard's famous consumer decision model (Howard, 1989), we examined decision making processes in online shopping. The target of this analysis was narrowed to online music services, which are the most popular online shopping service in Japan. Questionnaires about decision making for online music services were done with Japanese consumers.

2 DECISION MAKING PROCESS OF CONSUMERS

The Howard-Sheth buying behavior model presented in 1969 is one important model that explains consumer decision making. It explains the complexity of the consumer decision making process for incomplete information with an S-O-R paradigm that consists of three components: stimulus-organism-response. In this model, consumers are motivated by perceiving a stimulus, and then they shift to purchase after composing concepts to learn purchases. Furthermore, it is hypothesized that the problem is simplified as learning continues. As a result of experimental researches based on this model, Howard proposed the simple consumer decision model shown in Figure 1.

This simple consumer decision model is comprised of six interrelated components, as shown in Figure 1: information (F), brand recognition (B), attitude toward the brand (A), confidence in judging the brand (C), intention (I), and purchase (P). The three central components (B, A, and C) comprise the
buyer's brand image and the ABCs of consumer behavior.

“Information” (F) is what is received and stored. The perception is stored, not the stimulus. “Brand recognition” (B) reflects the extent to which the consumer knows enough about the brand to distinguish it from others. “Confidence” (C) is the consumer's degree of certainty that his/her evaluative judgment of a brand, whether favorable or unfavorable, is correct. “Attitude” (A), the second part of brand image, is the extent to which consumers expect the brand to satisfy their particular needs. “Intention” (I) is a mental state that reflects the consumer's plan to buy a specified number of units of a particular brand in a specified time period. “Purchase” (P) shows whether the consumer really does buy the brand. The influence from “information” (F) to “intention” (I) is connected by a dotted line. This arrow is added in routine problem solving cases when consumers repeatedly buy a low-price or a daily product.

![Howard’s consumer decision model modified for routine problem solving.](image)

### 3 ANALYSIS TARGET

#### 3.1 Online Music Service

With annual growth of about 2% since 2003, the size of the visual contents market was estimated in 2005 at approximately 11.3 trillion yen (about $0.1 trillion), broken down as follows: visual contents/5.3 trillion yen, audio contents/1.0 trillion, and text contents/5.0 trillion (White paper, 2007). Moreover, the music or game contents market has especially expanded. The usage rate of musical contents is 10.9%, which is the highest among such various contents as visual contents, audio contents, and text contents, of contents by Internet with PCs. Even using contents through mobile Internet, the usage rate of musical ring tones is the highest with 15.6% and the rate of downloading songs is 14.0%, which is second. Based on this present condition, we chose online music services as our research subject for online shopping through Internet.

#### 3.2 Survey Items

To adapt Howard's model to online music services, we considered the following survey items. Channels for acquiring information about online music services are considered communication among friends and families, information from such media as television, radio, and magazines, and such Internet sources as e-mails and web pages. “Brand recognition” corresponds to musicians and online music service providers. For “Confidence”, we considered confidence in artists and providers. “Attitude” items include sound quality, price, payment methods, data size, and outtakes or live versions. “Intention” included consumer's income, popularity, season, and advertising campaign. “Purchase” shows whether consumers really buy the music. This experiment examined the degree of experience with online music services.

### 4 QUESTIONNAIRE SURVEY

#### 4.1 Survey Condition

The questionnaires asked participants to indicate their degree of agreement with the above items based on a five-point scale: 5—strongly agree, 4—agree, 3—neutral, 2—disagree, and 1—strongly disagree. The question details are shown in the appendix. 282 consumers (164 university students, 106 high school students, and 12 others) answered the questionnaires.

<table>
<thead>
<tr>
<th>Download Medium</th>
<th>All</th>
<th>High School Students</th>
<th>University Students</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mobile phone</td>
<td>44.0%</td>
<td>64.2%</td>
<td>29.3%</td>
</tr>
<tr>
<td>PC</td>
<td>22.0%</td>
<td>9.4%</td>
<td>29.9%</td>
</tr>
<tr>
<td>Both</td>
<td>23.8%</td>
<td>19.8%</td>
<td>27.4%</td>
</tr>
<tr>
<td>Others</td>
<td>10.3%</td>
<td>6.6%</td>
<td>13.4%</td>
</tr>
</tbody>
</table>

Table 1 shows the ratios of download media. Many respondents use mobile phones. From Table 1, more than 80% of high school students and over 50% of university students use mobile phones, including respondents who use both PCs and mobile phones. Table 2 shows the numbers of download...
experiences from online music services. Almost 90% of students have downloaded music.

Table 2: Number of download experiences.

<table>
<thead>
<tr>
<th></th>
<th>All</th>
<th>High school students</th>
<th>University students</th>
</tr>
</thead>
<tbody>
<tr>
<td>51 ~</td>
<td>37.9%</td>
<td>38.7%</td>
<td>37.8%</td>
</tr>
<tr>
<td>11 ~ 50</td>
<td>25.2%</td>
<td>28.3%</td>
<td>23.8%</td>
</tr>
<tr>
<td>2 ~ 10</td>
<td>23.8%</td>
<td>25.5%</td>
<td>23.2%</td>
</tr>
<tr>
<td>1</td>
<td>2.5%</td>
<td>0.9%</td>
<td>1.2%</td>
</tr>
<tr>
<td>0</td>
<td>10.6%</td>
<td>6.6%</td>
<td>14.0%</td>
</tr>
</tbody>
</table>

4.2 Questionnaire Results

The questionnaire results are shown in Table 3. The symbols on the left row correspond to the number of the questions shown in the appendix.

The results of Information questions reveal that many students don’t think information by e-mail is very important. Furthermore, information by radio is not important because students do not usually listen to radio. Such information from Internet as web logs and web pages is relatively important. Direct word-of-mouth is more important than word-of-mouth by Internet. The most important information comes from TV; clearly, students are greatly influenced by TV.

As the results of “Brand recognition” questions, students pay much more attention to the musicians than delivery providers. Since consumers cannot directly examine the products of online shopping, it is often asserted that the reliability of the online stores is very important. However, with online music services, consumers can listen to trial songs as products, providing them a change to examine products almost directly. Therefore, consumers care about songs as products, but they have little concern about the reliability of providers.

The “Confidence” results for indirect purchases suggest that consumers do not have confidence that they will be satisfied with products and providers. Therefore, many respondents answered “Neutral.”

Regarding questions about “Attitude,” consumers are naturally interested in price. Trial services, sound quality, and data size are clearly important. Opinions are divided about accounting systems.

For questions about “Intention,” consumers want to repeatedly download and buy songs. Many consumers also want to download new releases or hit songs. Because the price of one song is cheap or because they can be billed for their purchases, the opportunity of income does not equal the opportunity of downloading.
4.3 Result Comparisons by Attributes

The questionnaire results were analyzed and compared based on respondent attributes.

(a) Download medium

We compared the results by three types of download medium attributes: mobile phones, PCs, and both. Significance of the compared results was found in “Information,” as shown in Table 4. Here, “positive” means the rate of respondents who answered “strongly agree” or “agree,” and “negative” means the rate of respondents who answered “disagree” or “strongly disagree.”

Consumers who use both mobile phones and PCs tend to consider word-of-mouth information from friends important. PC users tend to consider such word-of-mouth information from the Internet as web logs important. However, although information by the Internet is identical to web logs, such differences between mediums do not consist in information from web pages provided by corporations etc. Mobile phone users regard information from magazines and TV as important. Furthermore, no users obtained significant information from radio. Particularly, PC users regarded information from radio as unimportant.

(b) Generation

Table 5 shows the compared results with high school and university students. For “Information” questions, high school students are clearly more influenced by TV. Many university students gave negative opinions about magazines. On questions about radio, although answers from high school students equally included both positive and negative opinions, most university students had negative opinions about radio. Regarding questions about motivation in “Intention,” high school students gave many positive answers about downloading new releases or hit songs.

(c) Experiences

The compared results by experiences are shown in Table 6. “Heavy” shows the rate of users who have downloaded more than 51 times. In the same way, “Middle” is between 11 and 50 times and “Light” is less than 10 times. Experienced users believe that information from web logs, BBS, and word-of-mouth from friends are important. “Trial service” is regarded as important by all category users. But middle experienced users only consider it relatively important. The more experienced users also care more about data size. For questions about motivation in “Intention,” only seasonable songs are important for heavy users, though the other questions show no differences by experiences.

Table 4: Comparison by download medium.

<table>
<thead>
<tr>
<th></th>
<th>Mobile phone</th>
<th>PC</th>
<th>Both</th>
</tr>
</thead>
<tbody>
<tr>
<td>Positive</td>
<td>53.2%</td>
<td>40.3%</td>
<td>61.2%</td>
</tr>
<tr>
<td>F101</td>
<td>53.2%</td>
<td>46.8%</td>
<td>50.7%</td>
</tr>
<tr>
<td>F201</td>
<td>62.9%</td>
<td>43.5%</td>
<td>53.7%</td>
</tr>
<tr>
<td>F402</td>
<td>46.8%</td>
<td>21.0%</td>
<td>31.3%</td>
</tr>
<tr>
<td>F501</td>
<td>41.1%</td>
<td>25.8%</td>
<td>34.3%</td>
</tr>
<tr>
<td>F601</td>
<td>26.6%</td>
<td>14.5%</td>
<td>26.9%</td>
</tr>
<tr>
<td>F602</td>
<td>32.3%</td>
<td>16.1%</td>
<td>37.3%</td>
</tr>
<tr>
<td>F603</td>
<td>29.0%</td>
<td>16.1%</td>
<td>26.9%</td>
</tr>
<tr>
<td>F701</td>
<td>48.4%</td>
<td>40.3%</td>
<td>53.7%</td>
</tr>
<tr>
<td>F702</td>
<td>33.9%</td>
<td>41.9%</td>
<td>50.7%</td>
</tr>
</tbody>
</table>

Table 5: Comparison between high school and university students.

<table>
<thead>
<tr>
<th></th>
<th>High school</th>
<th>University</th>
</tr>
</thead>
<tbody>
<tr>
<td>Positive</td>
<td>60.4%</td>
<td>41.5%</td>
</tr>
<tr>
<td>F402</td>
<td>70.8%</td>
<td>34.3%</td>
</tr>
<tr>
<td>F501</td>
<td>50.0%</td>
<td>28.7%</td>
</tr>
<tr>
<td>F502</td>
<td>50.0%</td>
<td>28.7%</td>
</tr>
<tr>
<td>F601</td>
<td>30.2%</td>
<td>19.5%</td>
</tr>
<tr>
<td>F602</td>
<td>36.8%</td>
<td>25.0%</td>
</tr>
<tr>
<td>F603</td>
<td>37.7%</td>
<td>15.9%</td>
</tr>
<tr>
<td>F701</td>
<td>55.7%</td>
<td>25.6%</td>
</tr>
<tr>
<td>F702</td>
<td>73.6%</td>
<td>45.1%</td>
</tr>
<tr>
<td>F703</td>
<td>65.1%</td>
<td>38.4%</td>
</tr>
</tbody>
</table>

Table 6: Comparison by experience.
5 CONCLUSIONS

This paper described a survey of consumer decision making process for online shopping, especially online music services, the most popular service in Japan. Based on Howard's consumer decision model, questionnaires about decision making for online music services were done with Japanese 282 students. Differences between generations download mediums, and experiences were analyzed. As future work, we need more examinations with improving questionnaires etc.

REFERENCES

White paper, (2007). Information and Communications in Japan
http://www.johotsusintokei.soumu.go.jp/whitepaper/ja/h19/index.html
Chuanlan Liu, (2007). Modeling Consumer Adoption of The Internet as a Shopping Medium, Cambria Press.

APPENDIX

[Questions]

“Information”
- Direct word-of-mouth
  F101: Do you think word-of-mouth information from your friends is important?
- Word-of mouth by Internet
  F201: Do you think the information from web logs or BBS without registration systems is important?
  F202: Do you think the information from BBS with registration systems is important?
- E-mail
  F301: Do you think direct mail is important?
  F302: Do you think the information from e-mails for members only is important?
  F303: Do you think the information from fan clubs is important?
- TV
  F401: Do you think the information from TV commercials is important?
  F402: Do you want to get TV programs theme songs?
  F403: Do you think the information in musical TV programs is important?
  F404: Do you want to get movie theme songs?
- Magazines
  F501: Do you think the advertisements in magazines are important?
  F502: Do you think articles about musicians in magazines are important?
- Radio
  F601: Do you think the information from radio commercials is important?
  F602: Do you think the information from radio music programs is important?
  F603: Do you think information about musicians from special radio programs is important?
- Web pages
  F701: Do you think the information of hit charts on web pages is important?
  F702: Do you think the information from video delivery web sites is important?
  F703: Do you think the information in musician web sites is important?

“Brand recognition”
  B101: Do you try to learn a lot about music delivery providers?
  B102: Do you try to learn a lot about the musicians whose songs you are downloading?

“Confidence”
  C101: Do you have confidence in the reliability of delivery providers before purchasing?
  C102: Do you have confidence that you will be satisfied with a particular musician’s songs before downloading?
“Attitude”
- Trial
  A101: Do you think a trial service in delivery sites is important?
  A102: Even if you know the song, do you listen to a trial of it?
- Sound quality
  A201: Do you pay attention to sound quality?
  A202: Do you think such alternative versions as live recordings and outtakes are important?
  A203: Are you satisfied with the sound quality by which you can judge the song?
- Accounting system
  A301: Do you think price is important?
  A302: Do you think a monthly fee is a better accounting system?
  A303: Do you prefer payment per song as an accounting system?
- Data size
  A401: Do you consider the data size of songs?
“Intention”
- Service
  I101: Do you think the possibility of re-download is important?
  I102: Do you think the payment method is important?
- Money
  I201: When you have extra money, do you want to download a song soon?
  I202: On payday, do you want to download a song soon?
- Motivation
  I301: Do you want to download seasonable songs?
  I302: Do you want to download new releases?
  I303: Do you want to download hit songs?
  I304: Do you want to download award-winning songs?
- Campaign
  I401: If you can get one free song by purchasing of ten as part of a special campaign, do you want to download?
  I402: If you can get unlimited songs much for 300 yen (about $2.7) for a limited time, do you want to download?
TYPICAL PROBLEMS WITH DEVELOPING MOBILE APPLICATIONS FOR HEALTH CARE
Some Lessons Learned from Developing User-centered Mobile Applications in a Hospital Environment

Andreas Holzinger, Martin Höller, Marcus Bloice

Institute of Medical Informatics, Statistics and Documentation (IMI), Research Unit HCI4MED
Graz University Hospital, Auenbruggerplatz 2/V, Graz, Austria
andreas.holzinger@meduni-graz.at, mhoeller@sbox.tugraz.at, marcus.bloice@meduni-graz.at

Berndt Urlesberger
Department of Neonatology, Graz University Hospital, LKH-Universitätsklinikum, Austria
berndt.urlesberger@meduni-graz.at

Keywords: Mobile applications, touch screens, human-computer interaction, information systems.

Abstract: This paper provides an overview of the experiences gained during the design, development and implementation of mobile applications for use within the clinical domain. Current problems and issues that arose during the development of the software are documented and discussed. Medical professionals' opinions, both medical doctors and nurses, and their input were coupled with front end development (user interface design) and back end development (software engineering) to decide on the most optimum development path and to select the most appropriate environments. Most of all, this project can be seen as a further example that User-Centered Development (UCD) is necessary, however it is not sufficient when developing mobile, cross-platform, and future-proof applications for medicine and health care.

1 INTRODUCTION

During the past few years, improvements in the technology of touch screens, further miniaturization, lower power consumption, and longer battery life has made it possible to design and produce better mobile computers, tablet PCs, and small mobile devices such as PDAs or smart phones. The market for such mobile computing devices is rapidly expanding whilst at the same time the technological performance of these devices is steadily increasing, (Antinisca Di & Cecilia, 2007). In this paper we concentrate on discussing the lessons learned during the development of specific user-centered software for tablet PCs and define the tablet PC as our mobile computing device (Prey & Weaver, 2007). However, although mobile computers have been available for a relatively long time in hospitals (Forman & Zahorjan, 1994), different studies show that health care professionals are reluctant to use poorly designed mobile systems, as the patient care workload is very time constrained and can be extremely hectic (Brekka, 1995), (Holzinger & Errath, 2007). Mobile computer user satisfaction is certainly an issue and this has not been researched extensively (Ozok et al., 2008). Obviously, all aspects of Human–Computer Interaction (HCI) and Usability Engineering (UE) are of growing relevance, and must be especially considered when in the process of developing software for medical or health care purposes (Holzinger, 2007). This has resulted in a set of commonly accepted development practices, such as User-Centered Development (UCD) (Holzinger, Searle & Nischelwitzer, 2007), (Holzinger, Sammer & Hofmann-Wellenhof, 2006).

Although such considerations are important, we must accept that for an application to work well, all aspects must be taken into account, and there is the potential danger that HCI professionals and usability engineers tend to discount the underlying aspects of software engineering (SE) (Thimbleby, 2007).
2 MOBILITY IN HEALTHCARE

There are several areas of healthcare where mobile computers are necessary. On top of this, each area has its own unique requirements. In order to differentiate between them, we form two distinct sections: a) Round Assistance, which consists of help, or assistance, that a doctor might require when doing routine “rounds” of the wards to check the patients’ progress, etc.; and b) On-call Assistance, which consists of help or assistance that a doctor may need while on call-duty. These are not routine check-ups on patients; rather they are performed as and when deemed necessary by medical professionals who are on call-duty.

Two main clinical situations exist which determine whether or not a Patient Data Management System (PDMS) is used in the unit. Again, this depends on a number of factors. Wards that use such PDMSs face specific problems (Junger et al., 2001).

However, any pros of automatic data storage are also faced with cons: a) overloading the medical professional with data b) communication difficulties that occur with patient details. In both of the above cases, these problems will ensure that the future worth of mobile applications is secured.

All PDMSs have basic operation centers, which are generally PCs that are within a distance of two to three meters from the patient. There, the majority of the medical operations are performed (data observation, decision finding, drug prescription). Due to the huge amount of data available to the medical professional using the PDMS, it is difficult to gain a quick overview of the patient’s situation. Details can be found within the PDMS, but overviews are difficult to get. Switching between different patient details on the same computer/screen requires a lot of time and is actually not especially helpful.

One solution to this problem may be a set of two to three screens, on which data is displayed. Another solution would, of course, be the use of a mobile device, which communicates with the PDMS (via WLAN, for example).

A problem which is often encountered when doing rounds on the medical wards is that any discussion of a patient’s medical needs is often centered on the PC display belonging to the patient. Most often, one person sits while others stand around in a circle in order to view the display. This situation presents several problems, not least the difficulty in seeing smaller details on the screen (which does, of course, depend on the number of persons within group) and, furthermore, the situation is conducive to a communicative environment. An optimum solution to this problem would be to use mobile devices, which would aid communication and ease data visualization.

PDMSs are often located within intensive care units; units with standard care patients do not have PDMSs (this is due to the fact that in this case monitoring data does not have to be stored continuously). In such circumstances mobile applications could be used as information servers during a round (displaying laboratory data, etc.). Here mobile applications are essential.

In both the above scenarios the mobile device has to have a screen size and display resolution suitable for graphical data presentation. Therefore only tablet PCs or laptops are appropriate. Due to hospital budget shortages, duties-on-call have become more common. In such a situation a senior doctor is available by phone. Data presentation using mobile applications eases decision finding and is beneficial to the senior doctor. In such cases, mobile devices such as PDAs or smart phones may also be appropriate. However, this is true only in very specific circumstances.

3 METHODS AND MATERIALS

Within our project the prototype of a mobile system for visualizing a patient’s overall status during ward rounds was developed for the intensive care unit of the department of neonatology at Graz University Hospital, which is amongst the largest in Europe.

An automatic patient monitoring system stores a huge amount of various data at fixed intervals of 15 minutes for each patient. The measured data consists of vital signs, administered medications, expulsions, and so on.

The problem for the medical professionals was twofold: 1) The system’s user interface for viewing and analyzing the data is, however, very cluttered and containing extreme large amount of data in a unstructured way. Although a graphical timeline plot is provided, doctors often have to analyse the raw numeric data. The average time spent on the analysis of one patient is between 5 and 15 minutes, depending on the doctor’s experience. 2) The system is non mobile, requiring the medical professionals to proceed to the stationary PC’s.

The aim of our project was not only to provide mobility but also to significantly reduce the time spent on information perception, so the raw data only has to be analyzed if the patient is not in a good
condition. This is accomplished by providing instant visual feedback about the patient’s status using a combination of star plots and traffic light metaphor.

3.1 Device

The device used for the prototype was an LE1600 tablet PC by Motion Computing, which supports stylus and finger input. Extensive experiments about the differences between finger versus stylus input have already been undertaken (Holzinger et al., 2008b). Table 1 contains the technical specifications of this device.

Table 1: Technical specifications of the LE1600.

<table>
<thead>
<tr>
<th>Specification</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>CPU</td>
<td>Intel Pentium M at 1.6 GHz</td>
</tr>
<tr>
<td>Memory</td>
<td>1 GB</td>
</tr>
<tr>
<td>Display size</td>
<td>12.1” XGA LCD</td>
</tr>
<tr>
<td>Display dimensions</td>
<td>247 mm x 186 mm</td>
</tr>
<tr>
<td>Display resolution</td>
<td>1024 px x 768 px</td>
</tr>
<tr>
<td>Hard disk size</td>
<td>60 GB</td>
</tr>
<tr>
<td>Weight</td>
<td>1.4 kg</td>
</tr>
<tr>
<td>Physical dimensions</td>
<td>296 mm x 240 mm x 18.7 mm</td>
</tr>
</tbody>
</table>

3.2 Software

The software was written using Java 1.5 and the Swing user interface toolkit, where we had quite positive experiences from former projects (Holzinger et al., 2008a). Basically, services should be adapted at runtime to the features of the device. Also, end users should at any time specify that services are delivered to match certain parameters. For example, end users may request that an image be printed while specifying a particular resolution, format and/or number of colours (Stefano, Claudia & Luigi, 2007).

3.2.1 Visualization Front-end

The front end is based on so called “visualization modules”. Each module is created and configured by the medical doctor. It reflects a subset of the available data, containing interrelated values. The doctor chooses which values are contained within the module. The module “Circulation” for example would contain heart rate, blood pressure, oxygen saturation, etc. For each of these values the medical doctor sets three intervals for defining the alarm states of good, mediocre and bad. Each module has an overall alarm state, which can too be good, mediocre or bad. This overall alarm state is defined by the individual values’ alarm states.

If, for example, value A is bad, the overall alarm state of the module becomes bad too. In the visualization the alarm state is represented using the traffic light metaphor, i.e. the background of the star plot is filled with green, yellow or red.

3.2.2 Database Interface

The database in question was an Oracle 8i database, for which there are JDBC drivers available. Java Database Connectivity is an API (Application Programming Interface) which allows database-independent connectivity between the Java programming language and a wide range of databases, including Oracle 8i (Oracle, 2008).

Oracle have available a JDBC compliant driver for this database which allowed for relatively quick development of the interface between the Java program and the patient data.

Of course, speed was also an issue – without good response times the software would not be useful as a way of accessing data quickly and easily. However, in this regard JDBC and Oracle perform extremely well. Database and driver support, therefore, was not a technological issue when developing the application; however it was a factor which played a role in deciding which programming language to chose. It is also worth pointing out that were the JDBC drivers not available, this would have led to an extremely long development cycle and may not have been possible at all.

This is an absolutely crucial aspect that must be considered when developing mobile applications that access an external database: your programming language/platform/database combination must have:

a) the ability to perform the required task and, less obviously,

b) should have available quality drivers and libraries to ease development.

4 LESSONS LEARNED

4.1 General

As already mentioned in the introduction, mobile devices have increased in capability by many factors over the past few years, in terms of both features, such as integrated cameras, and raw processing power. Coupled with the fact that mobile devices now contain many of the attributes that constitutes a PC, several software development platforms have become available by various vendors, most notably
Google, Microsoft, and Sun Microsystems, among others. However, their usefulness in the area of medicine must be discussed.

Interestingly, the proliferation of mobile phones, and jointly, the popularity of games for mobile phones have cemented Java’s position as the environment of choice for developing mobile applications for small devices.

Having said this, developing applications in the area of medicine requires that a platform inhibits far different attributes than those platforms used to develops mobile games. In this section, we discuss the problems, and currently available solutions for developing real-world, cross-platform solutions for mobile devices.

4.2 Currently Available Cross Platform Development Environments

In the classical development world, if you wish to develop a piece of software you must usually first decide on the device on which you will develop your application. Palm, for example, offers an SDK for their devices which eases the development of applications. Therefore, software developed for Palm devices can only be run on Palm OS.

Limiting yourself to one single platform seriously diminishes your potential market when selling your software, or, in the case of the medical profession, may demand redevelopment (if, for example, a device or platform is no longer available when inventory is recycled). Cross-platform development, in many respects, eases this as you can develop software that runs on any device where the runtime environment is available (Bishop & Horspool, 2006).

Since Java’s philosophy of “write once, run anywhere” was incepted, several vendors have delivered cross platform languages.

Microsoft has developed .NET, and its Micro Edition competes with Java’s Mobile Edition in the mobile domain. A new player in this area is Google, who, along with 30 other technology companies, is currently touting the Android platform (Android, 2008). If a doctor or medical professional were to carry a device with them at all times, it would suffice to say that the device must just be portable. Tablet PCs have the advantage of having high system resources, large screens, and most run the ubiquitous Windows operating system.

On Windows, there are any number of cross-platform development environments that one could choose from, varying from the obvious, to the slightly more abstract yet equally capable alternatives (such as Adobe’s Air, and Microsoft’s Silverlight).

4.3 Java SDK

Creating game applications using the Java SDK seems to be the de facto way of producing applications for small devices, but here we must judge its worth as a platform for more serious application development.

A number of aspects of the Java SDK were analysed. For example, Java has been known to render fonts very poorly on screen, and has been a topic of discussion for quite some time, with various workarounds and techniques available to cure the problem. Another severe restriction is the missing support for floating point numbers in MIDP (Mobile Information Device Profile) versions prior to 2.0, though there exist third party workarounds for this problem. So long as the mobile device is MIDP 2 compliant, this is no longer an issue, however this is something which must be ascertained before development begins and could potentially be an issue in the future lifecycle of your software.

4.4 Alternatives to Java - AJAX

AJAX is a term used to describe a number of currently available technologies that when combined form a framework with which you can build desktop-like applications for the web (Turner & Wang, 2007).

The often touted examples are Google’s AJAX applications such as Gmail, Calendar or Maps – they allow drag & drop, ‘refreshless’ updating of information, and offer a desktop like and feel. Of course, being AJAX applications, they can run on any supported browser such as Firefox, Safari, or Internet Explorer, eliminating the need for the developer to worry about which operating system the user is running, as long as the operating system itself supports the browser. This has the added advantage that most of the high-end technology can reside on a server rather than on the user’s device. An AJAX application can therefore access an Oracle database, without the programmer having to worry about Oracle drivers being available for their framework/operating system combination.

One of AJAX’s advantages stems from the very fact that it was conceived as an internet platform – it is geared towards users implementing a point and click device rather than a keyboard. This bodes well for small devices, as input is generally carried out...
using a stylus rather than a keyboard. Using Google Maps on a traditional PC, for example, it is possible to find your street and house without having to input a single character into the keyboard (assuming some knowledge of geography). Creating applications, therefore, using AJAX geared towards mobile devices that utilize styli should demand no extra effort of the part of the programmer; in fact the inverse is true – it should be more instinctive to generate applications that do not require keyboard input.

4.5 Java Limitations

Java’s Runtime Environment (i.e. the Java Virtual Machine) is not platform independent. It is simply a runtime which is available on a (diminishing) number of platforms. Its source code, however, is platform independent, but this is also true of C/C++ code, often referenced as being platform specific. Consider the following example in Java:

```java
for (int i = 1; i <= 10; i++)
    System.out.println("Number: "+i);
```

And the following code written in C++.

```c++
for(int i = 1; i <= 10; i++)
    cout << "Number: " << i << endl;
```

Both these source codes are transferable between platforms. The C++ code requires that a compiler is available for the platform that you want to run your code on, but the same is true of Java – it requires that a Java VM is available on the machine you wish to run your code on.

The main difference is that the compiled output is not transferable, but the code is. Of course, programs are far more complicated than this, and even primitive types, such as integers, vary in size from C++ compiler to C++ compiler. And Java’s compiled output is transferable from one platform to another, without even the need to recompile, something impossible to achieve with a C++ compiled application. However, this still begs the question: is Java useful as a platform in the medical domain where usage on mobile or small devices is a must? The authors thinks so. While Java’s Virtual Machine is available only for Windows, Mac OS and Linux/UNIX, these are only operating systems available for tablet PCs as of the time of writing. It also seems very unlikely that another operating system will appear in the foreseeable future.

One more thing to consider, however, is that Java is not available for Windows Mobile or Palm (discontinued since the 12th of January, 2008). It is therefore the author’s opinion that AJAX could still be considered the most optimal solution in creating cross platform applications. However, it is unlikely that PDAs have the required resolution and screen size required to view patient information effectively. Of course, that is not to say AJAX cannot run on tablet PCs, this is certainly what AJAX was designed to do. Consider also the quickly changing medical field – AJAX applications when run, by definition, are always up to date (this, however, is also true of Java’s Webstart). Therefore, ruling out AJAX altogether would be foolish, it certainly has its niche, but perhaps not in the medical domain, or rather not in our specific area of patient care in the medical field.

4.6 .NET

Microsoft’s .NET framework is as platform independent as Java, in the sense that the runtime can be ported to any platform. Currently Microsoft only supplies a runtime for the Windows line of operating systems, but because Microsoft submits the specification for the Common Language Infrastructure to both ECMA and ISO, it is an open standard. Therefore, it is possible for third parties to create implementations of the framework on other platforms. This is currently the goal of the Mono project, which aims to port the framework to Linux. However, for our requirements in the medical domain, we required a far more concrete implementation, and Java officially releases several versions of its framework, something that ruled out .NET at this time for our purposes.

Again, inter-platform operability was an absolute requirement for us, as potentially many different machines and platforms would be using the software across the university hospital, and because we have no control over what devices the hospital purchases for its medical professionals.

4.7 Chosen Platform

The decision was made to opt for Java as the language of choice. The medical domain demands unique considerations that eventually ruled out most platforms that are currently available. Several years from now, .NET may be a contender, and AJAX was certainly a consideration. However, AJAX lacks the maturity and robustness required for the purposes of this project, and Java’s large library meant
development could be performed as rapidly as possible. It can be seen that the field of medicine demands far more considerations when developing a piece of software. Everything must be considered, from screen resolution, platform availability, speed, portability, library capability, and supported technologies. Java’ maturity, concrete standards (such as the JDBC API) and wide ranging third party support makes it the choice for medical software development.

5 CONCLUSIONS

It is clear that special considerations must be made when developing applications in the medical domain, especially if these applications should be platform independent, future proof, and mobile. There are a plethora of frameworks, environments, and programming languages available, each with their own specific advantages and disadvantages but only some are suitable for the medical domain. By reading this paper, it should be possible to save anyone a lot of research and work if they are considering writing a cross platform, portable application in the medical domain. Almost all considerations were taken into account, from screen resolution, doctors’ wishes, language suitability, and operating system capability. By working close to medical professionals, UI experts, and software engineers, it was possible to ascertain what special considerations must be taken into account when working in this field. By analysing these considerations, a number of concrete factors could be defined which eventually led to development path and programming environment that was chosen.

REFERENCES

Android (2008), Open Handset Alliance, Online available: www.openhandsetalliance.com/android_overview.htm l, last access: 2008-06-10
A STUDY OF FACTOR AFFECTING CUSTOMER SWITCHING BEHAVIOR OF MOBILE TELECOMMUNICATION 3.5G SERVICES

Jungwoo Lee, Eok Baek
Yonsei University, Republic of Korea
jlee@yonsei.ac.kr, baekkeok@yonsei.ac.kr

Sujung Sung
Autoeversystems, Republic of Korea
sue3649@yonsei.ac.kr

Keywords: Switching Behavior, Attractive of Alternatives, Network Externality, Switching Cost, Moderated Effect, 3.5G Mobile Telecommunication.

Abstract: Recently, Mobile Telecommunication businesses contend with each other to expand their customer base by using an aggressive marketing strategy. In order to determine if this strategy is effective, customer's and their switching behavior needs to be studied. This study identifies and analyzes direct, indirect factors affecting customer switching behavior such as attractiveness of alternatives, network externality, switching cost. The research results show that attractiveness of alternatives, network externalities have a direct impact on customer switching behavior. These factors also have a moderated effect on customer switching behavior but the switching cost does not influence customers directly.

1 INTRODUCTION

As the service industry becomes diverse and enhanced, competition increases among providers and brand preference is reduced (Fornell, 1992), prompting companies to practice defensive marketing in an attempt to maintain existing customers and attract new ones. In the mobile communication service industry, the cost for signing up a new customer is ten times greater than maintaining an existing client, thus customer retention has become an essential element of survival in the industry (Bolton, 1998). Accordingly, there is a need to examine customer behavior associated with switching to the competition. There are three main objectives of this study.

The first is to look into the dynamics of the mobile communication market based on previous studies conducted on the switching barrier.

The second is to explain how network externality is expressed in 3.5G services and exchanges influences with customer switching behavior.

Finally, this study attempts to conduct empirical analyses base on the correlations among the factors that affect switching behavior based on the two aforementioned theories.

2 LITERATURE REVIEW AND HYPOTHESES

2.1 Switching Behavior

Carpenter & Lehmann(1985) defined switching behavior as 'customer shift to other companies' products' and Reichheld & Sasser(1990) as 'customer defection'. Switching behavior can be explained as a concept that counters customer retention and intent to repurchase, and customer retention refers to the overall notion that encompasses the behavior of actual usage as well as the attitude toward the intent of continued usage (Oliver, 1999; Allen & Rao, 2000). Studies of customer retention in the service industry have focused on customer satisfaction and switching behavior as major factors that affect how customers can be retained. According to past studies, although the intensity varies in the relationship between
customer satisfaction and customer retention, the former is generally regarded as a contingent factor for the latter (Anderson and Sullivan, 1993; Jones, 1998; Oliver et al., 1993; Kim, Sang-Hyeon, 2002). This implies that high customer satisfaction has a negative effect on switching behavior, an opposing concept of customer retention and intent to repurchase. Accordingly, it can be concluded that the higher the degree of satisfaction about the current mobile communication service, the lower the probability of switching to 3.5G services. This study focuses on the switching behavior prompted by 3.5G services, and the definition of switching behavior shall include both churning to other service providers and upgrading to 3.5G services within the same provider.

**H 1: Customer satisfaction toward current mobile communication service has a negative(-) effect on switching behavior to 3.5G services.**

### 2.2 Switching Barrier

Jones(2000) defined switching barrier as the psychological and financial difficulties that a customer faces in the process of changing service providers. Switching barrier consists of the switching cost and alternative attractiveness, and these factors play moderating roles between customer satisfaction and switching behavior (Anderson et al., 1994; Jones et al., 2002).

Jones(1998) categorized the switching cost into continuity cost, contractual cost, searching cost, learning cost, initial cost and dissociation cost. However, since mobile communication providers offer similar services and usage of 3.5G services is not significantly different from conventional mobile phone, the searching and learning costs can be disregarded. Furthermore, since consumers are not involved in face-to-face interactions with their service providers, the dissociation cost of severing personal relationships does not have to be considered. Accordingly, this study only deals with continuity, contractual and initial costs.

Alternative attractiveness is defined as consumers' awareness of service providers that can replace current providers (Jones et al, 1998) and related to the concept of service augmentation in related literatures (Porter, 1980). Service augmentation refers to providing additional services to be distinguishable from other service providers. In turn, when there is an alternative that provides better services, consumers will disconnect the relationship with current providers and opt for the better option (Bradford, 2000). On the other hand, if the alternative lacks attractiveness, consumers will continue to use current services albeit not completely satisfactory (Porter, 1980).

**H 2: Switching barrier has effects in the relationship between customer satisfaction toward the current service and switching to 3.5G services.**

**H 2-1: Switching cost has a positive(+) effect between customer satisfaction and switching behavior.**

**H 2-2: Alternative attractiveness has a negative(-) effect between customer satisfaction and switching behaviour**

### 2.3 Network Externalities

First studied by Katz and Shapiro(1985), they defined network externalities as the concept that "the value consumer gains from a product or service is proportional to the size of the network that the product or service belongs to". In other words, network externalities come into consideration when the value of a product or service depends on the number of its users.

Consumers generally wish to be linked to large-scale networks because they expect products or services with large user bases provide higher quality and value. Accordingly, network externalities are referred to as consumption externalities of quantity, which does not provide much value on its own but signifies network characteristics of a product or service whose value increases with expansion of the user base (Shapiro, 1999).

The value of 3.5G services would be marginal without a significant user base of video telephony, which is the core of 3.5G services. Once the number of video telephony users increases, the value of 3.5G services will also increase due to the bandwagon effect. Thus the following hypotheses were established.
Table 1: Manipulative definition of variable.

<table>
<thead>
<tr>
<th>Category</th>
<th>Research Variables</th>
<th>Manipulative Definition</th>
<th>Items Evaluated</th>
<th>Questionnaire Structure</th>
</tr>
</thead>
<tbody>
<tr>
<td>Customer Satisfaction</td>
<td>Overall Customer Satisfaction</td>
<td>Customer recognition and sentiment after service experience</td>
<td>Overall level of satisfaction regarding fundamental services and provider</td>
<td>I-1, 2</td>
</tr>
<tr>
<td></td>
<td>Alternative Attractiveness</td>
<td>Degree of awareness regarding providers that offer alternative services</td>
<td>Whether 3G services are offered</td>
<td>II-1, 2, 3, 4</td>
</tr>
<tr>
<td></td>
<td>Continuity Cost</td>
<td>Cost associated with trouble in switching service providers</td>
<td>Psychological/financial cost associated with notification of new phone number</td>
<td>II-5, 6, 7, 8</td>
</tr>
<tr>
<td></td>
<td>Contractual Cost</td>
<td>Cost of losing financial benefits by staying with the current service provider</td>
<td>Cost of losing current service provider's membership services</td>
<td>II-9, 10</td>
</tr>
<tr>
<td></td>
<td>Initial Cost</td>
<td>Additional costs associated with switching to a new service provider</td>
<td>Sign-up fee for the new service provider</td>
<td>II-11, 12, 13</td>
</tr>
<tr>
<td></td>
<td>Network Externalities</td>
<td>Customer awareness of the user base for new services</td>
<td>Change in behavior according to network externalities</td>
<td>III-1, 2, 3</td>
</tr>
<tr>
<td></td>
<td>Switching Behavior</td>
<td>Customer's switching behavior from current service</td>
<td>Intent of continued use of services from the current provider</td>
<td>IV-1, 2, 3, 4</td>
</tr>
</tbody>
</table>

**H 4:** Network externalities have a negative(-) effect between customer satisfaction toward current service and switching behavior to 3.5G services.

**H 5:** Network externalities have a positive(+) effect in switching behavior to 3.5G services.

3 RESEARCH METHODS

In order to verify the hypotheses of this study, manipulative definition was established for each variable based on the previous studies described in Chapter 2 to elicit evaluation items for this study. Online surveys were conducted from November 12 to 16, 2007 for male and female mobile telephone users between the ages 10 and 60. Among 180 questionnaires retrieved, those with incomplete or insincere responses were discarded and the remaining 158 were used as the survey data for the study. Likert 7-point scale was used for every question except for the items regarding demographics, and SPSS 12.0 was used to perform statistics analysis, reliability analysis, factor analysis and regression analysis.

4 DATA ANALYSIS

4.1 Characteristics of Sample Space

The statistics related to the 158 respondents for the survey. The male-female ratio was somewhat
balanced at 53.2% and 46.8%, but 83.8% of the respondents were in the 20~40 age group.

4.2 Analyses of Variable Scale Reliability and Propriety

The internal consistency method was used for verifying reliability of this study. The survey questions were used without modification because there was no indication of reliability being undermined, and the Cronbach's Alpha coefficient for the factor variables used in the study were between 0.6 and 0.9, displaying a high level of internal consistency.

In addition, factor analysis was performed for each item that constitutes a variable to analyze propriety and find out whether each variable was isolated and assessing a unique property. VARIMAX rotation based on major element analysis was used to elicit factors, and factor loading was set at 0.4 for those with Eigen value of 1.0 or higher, which yielded 5 factors. The analytical result indicated that the two variables that explain switching cost - contractual cost and initial cost - should be grouped as a single factor. It was determined that the respondents regarded both contractual and initial costs as being associated with switching to or signing up for a new service provider, and the two variables were redefined as the initial switching cost.

4.3 Verification and Analysis of Hypotheses

Verification for the 5 research hypotheses were conducted in two stages. First, multiple regression analysis was used to verify the factors that directly affect switching behavior. H1, H3 and H5 are classified as direct factors. The second stage involved analyzing the moderating effect that influence the relationship between current mobile service satisfaction and switching behavior, for which H2 and H4 correspond with. Hierarchical multiple regression analysis was used to analyze the moderating effect. In addition, the average value of each factor was calculated for regression analysis. Adoption criteria of hypothesis were set at 95% reliability and a significance level of 0.05 or lower.

Summarizing the analytical result, 5 of the 7 hypotheses established for the study were adopted and the remaining 2 were discarded.

### Table 2: Result of Analyses.

<table>
<thead>
<tr>
<th>Hypotheses</th>
<th>Result</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hypothesis 1</td>
<td>Adoption</td>
</tr>
<tr>
<td>Hypothesis 2</td>
<td>Partial adoption</td>
</tr>
<tr>
<td>Hypothesis 2-1</td>
<td>Discard</td>
</tr>
<tr>
<td>Hypothesis 2-2</td>
<td>Adoption</td>
</tr>
<tr>
<td>Hypothesis 3</td>
<td>Partial adoption</td>
</tr>
<tr>
<td>Hypothesis 3-1</td>
<td>Discard</td>
</tr>
<tr>
<td>Hypothesis 3-2</td>
<td>Adoption</td>
</tr>
<tr>
<td>Hypothesis 4</td>
<td>Adoption</td>
</tr>
<tr>
<td>Hypothesis 5</td>
<td>Adoption</td>
</tr>
</tbody>
</table>

4.4 Additional Analysis: Effect of Intensifying the Switching Cost Research Model

Additional analysis was performed for this study to assess how switching costs customer satisfaction and switching behavior based on the fact that the initial switching cost in particular has influence on the two factors. The respondents were divided into two groups based on whether they felt the initial switching cost was high or low to analyze the effects between degree of satisfaction and switching behavior. The result indicated that there is significant influence between level of satisfaction and switching behavior for the group that felt the initial switching cost was high, but the influence insignificant for the other group.

The result corresponds with the basic research model demonstrating that customers who feel that there is a high cost of switching mobile service providers recognize the cost as a switching barrier and will not display switching behavior. Therefore, it can be learned from the additional analysis of this
section that the initial switching cost reinforces the research model of this study.

5 CONCLUSIONS

This study analyzed a comprehensive model of switching behavior to examine consumer awareness of the new 3.5G mobile communication services and the factors that affect the switching behavior to 3.5G services. The study was conducted in two stages. First, the study assessed the factors that directly affect the level of satisfaction as well as network externalities, perceived switching cost and alternative attractiveness of 3.5G services. Second, the interaction effects of network externalities, alternative attractiveness and switching cost were examined as moderating variables between satisfaction and switching behavior.

The expectation of 3.5G services providing various distinguishing features from conventional mobile communication to offer new sources of satisfaction has been reflected in this study. In the case of video telephony, the flagship 3.5G service, network externalities significantly affect consumers' switching behavior in terms of the perception of how many others were using the service.

As for the switching cost, number portability and sign-up fee waiver offered by the service providers allow consumers to alleviate the burden of notifying their new numbers to their contacts or signing up for new services.

Whereas previous studies on the subject have focused solely on the switching barrier, this study bears significance in that network externalities were analyzed from the consumer behavior perspective. Moreover, the study provides a foundation for establishing a practical strategy for securing a 3.5G service customer base to gain a competitive edge.

However, this study failed to cover factors such as consumer usage characteristics and more than 80% of the survey respondents were in the age group between 20 and 40. It is therefore necessary to diversify the age groups of the sample space to increase the accuracy of the study. Questions have been raised regarding the effectiveness of the strategies implemented by the mobile communication providers to maintain and acquire customers for 3.5G services, but this study did not examine whether customers display switching behavior according to the strategies. In turn, it would be necessary in future studies to assess whether service providers' strategies are effective based on the mechanisms explained in this study as well as the number of 3.5G service subscribers and changes in the actual profit structure.

REFERENCES

Bradford, Neeru Sharma, Patterson G., 2000, Switching costs, alternativeness and experience as moderators of relationship commitment in professional, consumer services, International Journal of Service Industry Management, 11
ADOPTION OF NEAR FIELD COMMUNICATION TECHNOLOGY IN BUSINESS TO CONSUMER SERVICES

Arto Wallin, Juha Häikiö and Jaana Määttä
VTT Technical Research Center of Finland, My, Kaitoväylä 1, Oulu, Finland
Arto.Wallin@vtt.fi, Juha.Haikio@vtt.fi, Jaana.Maatta@vtt.fi

Keywords: Adoption of mobile services, Near Field Communication, NFC, NFC-based mobile services.

Abstract: Touch based interaction has been found out to provide fast and easy way to use mobile services. Near Field Communication technology (NFC) is one of the most promising ways to conduct touch based interaction with mobile phones. In this paper our aim is to identify challenges of the adoption of NFC related business to consumer services. The results of this study are based on three pilot cases conducted in Finland in 2007. Our research presents findings based on qualitative company interviews and both quantitative and qualitative user questionnaires. We conclude that there are several challenges in the adoption of NFC-based applications. However, expect for the shortage of NFC enabled mobile phones, challenges of the adoption may be reduced with careful planning and implementation of NFC-based applications.

1 INTRODUCTION

Although the diffusion of mobile phones has been outstanding during last years, use of mobile services has not gained comparable success. One reason behind the slow diffusion of mobile services might be that the usability of mobile services has not been at a sufficient level. This is partly caused by the poor user interfaces of today’s mobile phones, which can be experienced when, for example, trying to type text or navigate in menus using mobile phone’s keypad. Difficult user interfaces may even totally prevent some users, such as elderly people, from using mobile services.

Difficult user interaction of traditional mobile services has created a need to develop easier and more intuitive user interfaces. A promising solution for this problem is based on a physical selection method, in which users interact with the digital environment by touching Radio Frequency Identification (RFID) tags with the mobile devices (Välkkynen, Niemelä & Tuomisto, 2006). According to the previous study by Riekki et al., services used by touching were experienced as easy to use and learn. Touching was also appreciated, because “it gives clear feeling of control” (Riekki, Salminen & Alakärppä, 2006).

Touch based user interaction can be conducted with a mobile phone by using new short range wireless communication technology called Near Field Communication (NFC). One of the main advantage of the NFC technology is that it can be used to communicate with other mobile devices or tags without a configuration (Calvet, 2005). Therefore, the communication with the NFC enabled mobile phone should be as simple as touching objects.

In addition to the advantages for consumers in form of usability, use of touch based interaction technology may provide business benefits for service providers. In the case studies reported in this paper, NFC technology was used by the service providers with an aim to improve the effectiveness of their service processes. Hence, besides increasing customer satisfaction, the service providers expected that the NFC technology might also provide indirect cost savings to them.

Despite the clear benefits of a touch based user interaction, the adoption of NFC has not been as fast and wide as some have expected (Balaban, 2007). After the fade of initial hype, there still exist some challenges that have to be overcome before the wide scale adoption of NFC is realized. This paper aims to identify and analyse these challenges using both quantitative and qualitative data obtained from three pilot cases conducted in the city of Oulu during the autumn 2007. Data contains both the service providers’ and the consumers’ perceptions about the challenges of adoption.
The paper is organised as follows. Section two introduces a theoretical background and earlier studies related to the adoption of NFC technology and mobile services. In the section three, we present our research methodology and pilot cases. We will then present the results of our empirical studies in the section four, which is followed by the discussion and conclusion.

2 RELATED RESEARCH

Theory and experiences about the adoption and diffusion of innovations has been studied for years. Even today many papers are based on diffusion innovation model introduced in 1962 and later refined by Rogers (1995). Innovation adoption theories focus on identifying and describing factors that have some effect to the adoption. Kargin and Basoglu (2006), for example, introduce framework for adoption factors of mobile services. They state that usefulness of the mobile service and past experience influence on the consumer’s attitude towards new technology and thus to the adoption of technology.

Although there is plenty of research about the adoption of mobile services (e.g. Nysveen et al. 2005; Bouwman et al., 2007; Carlsson et al., 2005), the research about the adoption of NFC-based mobile services is almost non-existent. Main reason for this is the novelty of NFC technology. Despite of the novelty, touching as an interaction technique in NFC-based applications has been studied in several studies. These studies consider touching with a mobile phone as a very promising interaction technique (Rukzio et al., 2006; Riekki et al., 2006; Välkkynen et al., 2006; Anokwa et al., 2007).

Some studies have already evaluated usability of the NFC technology in form of case studies. Jaring et al. (2007) survey the results of six small NFC pilots, which are focused on improving mobile solution workflows and usability. They conclude that NFC based solutions are easy to use, but the small and limited keypad poses challenges to the design of NFC applications. Häikiö et al. (2007) have studied the use of the NFC technology in elderly care environment. The study provides evidence that the adoption of the touch based interaction does not require a specific knowledge or experience with technology. In addition, elderly people, who had impaired motor skills, were able to use a touch-based user interface in the pilot test.

User acceptance of services is a crucial condition for the success of mobile services. Some existing acceptance models emerge issues, which can be used as an evaluation framework for mobile services. Kaasinen (2005) has developed Technology Acceptance Model for Mobile Services. This model has its roots in Technology Acceptance Model created by Davis (1989). In Kaasinen’s model user acceptance of mobile services is consisted of three factors, which affect the users’ intention to use a mobile service. These factors are perceived value of service, perceived ease of use and trust. According to Kaasinen (2005), the fourth factor, ease of adoption becomes more central, as a user proceeds from intention to the actual use. These fore-mentioned factors of Technology Acceptance Model for Mobile Services can be seen also as a foundation for examination of NFC-based mobile services.

3 RESEARCH SETTING

The purpose of our research was to examine suitability of the NFC technology to selected business to customer services and to find out general barriers to adoption of NFC technology. Since this topic is quite new, our research approach is explanatory and it presents finding of both qualitative and quantitative research conducted in three individual pilot cases. Table 1 summarises the characteristics of these pilots including the user amount, the duration, the user segment and the goal of the pilot. More detailed information about the pilots is presented in the following subsections.

The objective of the qualitative research was to find out detailed information about business actors’ expectations, experiences and challenges considering the adoption of NFC in their service processes. Interviewed actors included upper level managers from two private companies and two public organisations. At least one interview was conducted in each of these four organisations. In addition, restaurant pilot included three interviews with operational level employees. All interviews were conducted after the finish of the pilots in the beginning of 2008. Interviews followed semi-structured format and lasted from one to two hours each. The interviews were recorded and transcribed for the further analysis.
### Table 1: Summary of the analysed pilots.

<table>
<thead>
<tr>
<th>Pilot #</th>
<th>Pilot name</th>
<th>Responses / Users</th>
<th>Duration</th>
<th>User segment</th>
<th>Goal of the pilot</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>SmartParking</td>
<td>48/51 + 5/5 traffic wardens</td>
<td>2 months</td>
<td>Frequent parkers/traffic wardens</td>
<td>Improve effectiveness of parking in city area. Provide fast, easy and efficient parking solution for customers.</td>
</tr>
<tr>
<td>2</td>
<td>Fast track of restaurant Pannu</td>
<td>23/27</td>
<td>5 weeks</td>
<td>Restaurant’s regular customers.</td>
<td>Provide fast ordering and payment for busy lunch time customers.</td>
</tr>
<tr>
<td>3</td>
<td>Oulu city theatre</td>
<td>101/141</td>
<td>2 months (8 plays)</td>
<td>Employees from several different companies.</td>
<td>Improve customer experience and provide fast and easy ticket checking and refreshment reservation.</td>
</tr>
</tbody>
</table>

Companies selected to the pilots were expected to benefit from the use of NFC technology in their business processes. In order to find out a clearer vision for the NFC technology opportunities, the first section of the interview concentrated on organisation’s background information, a position at the market and a current technology usage. The second part surveyed the effects of NFC technology, which were detected during the pilots. The aim was to find out what benefits and challenges are directed to the business actors’ service processes. The last part considered pros and cons of the NFC technology in a more general level.

The objective of the quantitative research was to find a general overview about the drivers and barriers for the customer adoption of NFC in different services. The data was collected with web and paper based user questionnaires. Pilot participants received an e-mail with a link forwarding them to the questionnaire or they filled the paper questionnaire after a pilot period. The number of customers participating in the pilots is reported in the table 1. Altogether, there were 224 pilot participants who used NFC enhanced services in the pilots. Final response rates of questionnaires of pilot 1, 2 and 3 were 94%, 85% and 72%, respectively. The response rates can be considered to be very high especially in pilots 1 and 2.

All pilot participants received Nokia 6131 NFC phone for the pilot period. These NFC-enabled phones were equipped with proper applications for the pilot use. NFC related services and information was available without a charge to the individual pilot customers. Moreover, pilot participants were allowed to use NFC enabled phones for free of charge during the pilots, also for their own purposes.

### 3.1 Pilot 1: SmartParking

The objective of the pilot was twofold. Firstly, the aim was to provide flexible and efficient mobile parking solutions to the citizens. Secondly, traffic wardens tested a new NFC based parking control solution, which was envisaged to improve the effectiveness of the parking control. Although the NFC based parking control was highly appreciated by the traffic wardens, it will not be analysed in detail, since the focus of this paper is on the business to customer services.

A main pilot group consisted of 51 customers who parked frequently at the city area. Customers were able to use SmartParking for on-street parking and also in one parking hall located in the city centre. On-street parking transaction was initiated by touching an NFC tag attached to car’s windscreen. A user had to then select a parking zone either by touching one of the parking zone tags attached to fare collection machines and lamp posts, or alternatively, selecting a zone code manually from the phones menu. The selection of the parking zone by touching is illustrated in Figure 1. Parking time started running after the touch to the parking zone tag and it ran until user came back to her/his car and touched again the tag on the windscreen. Hence, SmartParking allowed charging from the exact time the car was parked.

![Figure 1: A user selects a parking zone by touching an NFC tag on the fare collection machine.](image-url)
an NFC reader near the gate. The touch to the NFC reader opened the gate and parking time started running. Customers repeated the same operation when they left the parking hall, which stopped the parking time. In addition to the NFC-based parking application, customers had a possibility to download information to their phones by touching info tags, which were attached on parking meters.

3.2 Pilot 2: Fast Track of Restaurant Pannu

An initiating problem behind the pilot was that busy workers want to have fast service at the restaurant because of their short lunch break. However, since most employees come to the restaurant at the same time, traditional service, where a waiter asks for orders, may be quite slow. In order to find a solution for this problem, a new “fast track” service concept was developed. The fast track provides an opportunity for the busy customers to select their lunch by touching the RFID tagged menu with their NFC enabled mobile phones, which is illustrated in Figure 2. The lunch order was directed straight to the kitchen without intervention of the waiter. In addition to placement of lunch order, customers were also able to pay for their lunch using NFC phones and information tags provided a possibility to download information to the mobile phone.

3.3 Pilot 3: Oulu City Theatre

The target of the pilot was to develop theatre’s service processes and deliver more added value to theatre customers with more flexible and easier NFC-based activities. Since the queuing and slow ticket checking activities were seen as bottlenecks, one of the main activities in this pilot was a mobile phone specific ticket stored in the back-end system. The ticket was checked before entering to the play by ticket checkers’ NFC enabled mobile phone. In practice, a theatre-goer and a ticket checker touched each others’ mobile phones. Theatre pilot included also other NFC related activities, such as, pre-reservation for interval refreshments, reclaiming of a programme for the play, downloading information via info tags and downloading video trailers from smart posters.

The pilot participants, 141 persons with average age of 44.5 years, tested and used NFC based services before and during the plays. Each participant visited at least in one play. The participants were informed about the theatre pilot by sending related information to companies locating in Oulu region. Selected participants were groups of employees from diverse industry areas. After the participants had visited the theatre and watched the play, they received web questionnaire to give a feedback about the use of NFC in theatre environment.

4 RESULTS FROM THE PILOTS

Analysis of the data obtained from the pilots reveals several common challenges for the adoption of NFC based services. In the following these challenges are divided into two categories based on the actor whom the challenge primarily concerns.

4.1 Customers’ Barriers to Adoption

The use of info tags revealed to be easy regardless of use experience concerning NFC-enabled mobile phone. Use period of NFC-enabled mobile phone was very short in the theatre pilot when compared to other pilots. However, users experienced touching also for the payment of their lunches. The payment transaction was conducted by tapping the NFC enabled phone on the contactless point of sales terminal. The used terminal was a commercially available RFID based terminal provided by a cash register provider.
and learning to use the info tags very similar in the theatre pilot and the restaurant pilot. A following table presents average values concerning experience about learning to use info tags and successful touching. The scales in questionnaires were from 1 ("very difficult") to 5 ("very easy") concerning learning and successful touch and from 1 ("totally useless") to 5 ("very useful") concerning usefulness of info tags’ content.

Table 2: Average values concerning experiences about learning, touching and usefulness of tags’ content pilot-specifically.

<table>
<thead>
<tr>
<th>Pilot/Feature</th>
<th>Theatre Pilot</th>
<th>Restaurant Pilot</th>
</tr>
</thead>
<tbody>
<tr>
<td>Learning</td>
<td>4.59</td>
<td>4.64</td>
</tr>
<tr>
<td>Successful touch</td>
<td>4.63</td>
<td>4.60</td>
</tr>
<tr>
<td>Usefulness of info tag content</td>
<td>3.00</td>
<td>3.07</td>
</tr>
</tbody>
</table>

Opinions about usefulness of info tags’ content varied among pilot users. 14 out of 23 users used info tags in the restaurant pilot. Info tags were touched 143 times during the pilot. Thus, there were approximately 5 touches to info tags per user in the pilot. Those, who did not use info tags in restaurant, stated that they did not have need to use those tags or content of tags was not interesting. Three of those users, who used info tags, saw that info tag content was very useful for them. However, many users did not see available info tag content very useful for themselves in lunch time.

Set of info tags was also available for pilot users in the theatre pilot. Experience about usefulness of info tags’ content was very similar with experiences in the restaurant pilot as shown in the table 2. Some users saw that info tags were nice extra to the theatre experience. In contrast, some users stated that info tags did not provide interesting information content for them. Info tags were touched 275 times in the theatre pilot. This means approximately 2 touches to info tags per user during the pilot. Smart poster tags, which offered information about theatre plays, were more popular. They were touched 345 times during a pilot period. Most of those users, who did not use smart posters, told they did not have time to use poster tags or they did not even notice these tags in the theatre.

The placement of info tag application in SmartParking pilot differed from the tag placement in two other pilots. The info tags were placed outside where those were attached to the parking meters (Figure 3). Context of use was clearly different than in other pilots, because car drivers wanted usually to get through the parking payments very quickly. Therefore, the other services in a parking meter were easily ignored. Moreover, many of those users, who used info tags, did not see the content of tags necessary for themselves during parking. For these reasons, the use of info tags was quite low in the SmartParking pilot.

Figure 3: Info tag menu in a parking meter.

None of the pilot users had ever used an NFC-enabled mobile phone before the pilot tests. Thus, in order to use NFC-based mobile services after the pilot period they should invest in a new NFC-enabled mobile phone. Although many users were willing to use NFC-enabled mobile services also in the future, a demand for buying a new NFC phone was experienced as a confine to the future use. Furthermore, because of the limited supply of NFC services, many persons were not ready to buy a new NFC-enabled mobile phone just for few services.

In the theatre pilot many users saw a price of services as an essential criterion for the future use of services in theatre environment. In the restaurant pilot 6 out of 22 users would be ready to pay for the services available during the pilot. Some users stated that they would use these services also in the future, if the service would be inexpensive. Overall, many users required that services should provide a clear benefit for them before they would be ready to pay for the services.

4.2 Service Providers’ Barriers to Adoption

Service providers’ challenges and barriers to the adoption are based on the interviews with business actors. Interviewed actors emphasized the fact that
benefits of the NFC technology are not expected to be realized immediately. However, they all believed that the NFC technology might somehow benefit their business in the future.

Firstly, interviewees experienced a low penetration rate of NFC enabled mobile phones to be main obstacles in short term adoption of NFC technology. The penetration rate was considered to be especially important in the cases of SmartParking and city theatre, which require large customer bases to be justifiable from the business perspective.

Secondly, substitutive technological solutions were considered as a notable challenge to the NFC adoption in piloted use cases. This challenge was highlighted in the discussions with managers of restaurant and parking operator. According to the restaurant manager’s opinion, the need of NFC based ordering application may be questioned when waiters’ wireless handheld devices become more general. However, interviewee considered NFC to be very potential complementary payment technology to be used in restaurants. From the viewpoint of parking operator, some other technological solutions, such as long distance RFID or licence plate recognition, are competing in the same category with NFC when investment decisions are made. Thus, interviewee from parking operator stressed that the new technology should bring some additional value in comparison with other potential technologies.

Lastly, interviewees regarded that the interoperability of NFC with existing systems is of considerable importance. Municipality’s parking supervisor stated that it would be very difficult to start using new parking solution that would be incompatible with existing mobile parking systems. Importance of interoperability was also noticed in restaurant environment. This was stressed in the interview with the restaurant’s management, since some of the value that customer perceived from the mobile payments was eroded because of the integration problems of the NFC point-of-sales reader and the cash register.

5 DISCUSSIONS

In this section we propose a sequence of research propositions related to the barriers to adoption of NFC technology and suggestions how these barriers can be lowered or even totally removed. Following propositions are provided with illustrations and explanations from previously introduced pilot cases.

**Proposition 1.** Real added value is a required for the customers’ adoption of NFC enabled services, especially if customers have to pay for it.

As the results presented in the section 4 point out consumers require meaningful content in the information tags, which provides added value for them. Furthermore, a content of info tags should be suitable with environment and the context of use. Significance of the environment was emphasized in the theatre pilot, as theatre play related information was more interesting than a content, which was not directly related to the theatre visit. In the SmartParking pilot car drivers tended to use a keypad of a mobile phone more frequently than zone tags in a parking zone selection. It is rather easy to remember the different parking zones in the city centre of a relatively small city, thus the perceived value of the NFC application can decrease in specific environment more than in another.

More meaningful services and content might also increase customers’ willingness to pay for the services. As restaurant and theatre pilots pointed out, willingness to pay for the service and content varies clearly between customers. Many customers were not even willing to pay at all for the service or content available in the pilots. More personalised services might be one solution for this problem, since those could deliver more focused content to the customers.

To conclude, before an implementation of an NFC enabled mobile service, it would be essential to identify information and service contents, which can provide added value to the customer in specific environments. This is even more important in cases where service provider requires that larger customer groups would be ready to pay for the NFC-based services.

**Proposition 2.** Penetration rate of the NFC enabled phones has to be substantially higher before a business to customer service intended for large customer groups can be profitable.

Although there are several niche solutions where NFC technology can be currently used successfully, large scale business to customer services are still waiting to be rolled out. The main reason for the lack of large scale NFC enabled customer services is the same as the interviews with companies and public sector organisation pointed out: NFC services intended for the large customer groups require a wider diffusion of NFC enabled mobile phones. Therefore, the major challenge is how to get NFC enabled mobile phones to the users who do not even necessarily want a new phone. In addition, currently
the shortage of different models of NFC-enabled mobile phones limits users’ possibilities to select a phone, which responds to their needs. The NFC phone model selection is expected to increase in the near future as different manufacturers have published plans to introduce new NFC enabled mobile phone models. So, at the moment, it seems that this challenge might be gradually fading.

**Proposition 3.** Taking into use of NFC services should be made as easy as possible for a customer.

A basic idea behind the NFC enabled mobile services is that they should be easy to use and adoption of use should not be difficult. Hence, in order that the adoption of a new NFC-based service would be as fluent as possible, it should not require too many preliminary operations from a user before the actual use of the service. Different initial preparations, such as registration and requirement to install an application, reduce consumers’ interest on using the service. At worst, those can completely prevent the adoption of an application.

In the pilots of this study, users received pre-installed and configured NFC enabled phones for the pilot use. Therefore, users were able start the use of phone immediately without time-consuming preliminary operations. In commercial use the pre-installation of all NFC related application to the mobile phone is not likely to be feasible. One possibility to solve this problem is to offer settings and configurations into a mobile phone by touching an NFC tag. In that way the easiness of taking NFC services in to use might be ensured.

**Proposition 4.** NFC service should be made interoperability with an existing technological infrastructure.

Fourth proposition is very important in the cases where NFC based service is taken into use alongside with other information systems. The importance of a fluent integration is emphasized, because it may be very expensive and complex for the service providers to maintain two concurrent technological solutions. Major issues may arise, for example, if the service provider of an existing information system is not willing or capable to integrate NFC based services to the existing system. This may result in high costs, if the whole information system has to be redesigned or it may even totally prevent a company from adopting a new NFC based service.

**Proposition 5.** Business value of NFC technology should be more clearly represented.

A field of research is quite unanimous that the measurement of benefits of mobile technology is difficult and the assessment of potential future benefits is even more difficult (e.g. Heijden & Valiente, 2002). In spite of that it may be critical from the service provider’s viewpoint to be able to assess the business value of an investment to new technology. Image benefits of being in the forefront of technology developers may be counted as a benefit by some business actors. However, when companies from more traditional industries plan and compare the investment on NFC technology to other technologies, pointing out the business value is essential. Therefore, more effort should be put on the assessing the total benefits of the NFC technology.

As opposed to the benefit assessment, more effort should be also set to assessment of the total costs of the NFC technology. Many indirect costs, such as management of the tags or protecting the confidentiality, integrity and availability of the services should be remembered when planning to implement touch based services for customer markets. These issues may turn out to be of significant importance if NFC solutions are used in public places where those are left without the supervision and those may be exposed to vandalism or difficult weather conditions.

6 CONCLUSIONS

NFC technology makes it possible to implement mobile services and applications, which are easy to learn because of intuitive and natural interaction technique. Although NFC technology has already been successfully adopted in some business to business services, there are still some challenges to be overcome before the wide scale adoption of NFC in business to consumer services can be reached.

Based on the results of three pilot cases, which were analysed in this study, the main barrier to the short term adoption is the shortage of NFC enabled mobile phones. Also other challenges were identified, however, these challenges can be taken into account at the level of individual company when NFC enabled services are planned and implemented. Therefore, these issues should be carefully considered before the implementation phase of the service.

Our research has some limitations, which leaves place to further research. Firstly, NFC services were provided free of charge to the customers during the pilots, which may not necessarily be the case in real
life. Although customers were asked about the willingness to pay for the services, more realistic long term pilot should be conducted to find out real business potential of the services. Secondly, the selection of pilot participants was mainly based on the customers’ own eagerness. For example in the SmartParking case, the user segment of pilot customers was mainly restricted to frequent parkers. However, it would be important to know how customers who need parking services less frequently experience the service. Hence, future research should pay focus on different customer segments and their willingness to adopt NFC services.

ACKNOWLEDGEMENTS

This work was done in the SmartTouch (www.smarttouch.org) project (ITEA 05024), which is a project within ITEA 2 (Information Technology for European Advancement), a EUREKA strategic cluster programme. The SmartTouch project has been partly funded by Tekes, the Finnish Funding Agency for Technology and Innovation.

REFERENCES


Calvet, J.C., 2005 The Role of RFID in the Mobile Phone. Teletronik 3(4).


TAXONOMY FOR MOBILE TERMINALS
A Selective Classification Scheme

Gunther Schiefer and Michael Decker
Institute of Applied Informatics and Formal Description Method, University Karlsruhe (TH), 76128 Karlsruhe, Germany
Schiefer@aifb.uni-karlsruhe.de, Decker@aifb.uni-karlsruhe.de

Keywords: Simple Phone, Feature Phone, Smartphone, Handheld, Web-Tablet, Netbook, Mobile PC, Tablet-PC, classification, mobile Terminal.

Abstract: There is a great number of different types of mobile computing devices like cellular phones, Handhelds and notebooks. However in literature so far there is no common accepted definition for these terminal types. Especially the term “Smartphone” is used for very different types of mobile phones. In our article we therefore propose a comprehensive taxonomy for mobile terminals that clearly distinguishes between different classes of terminals.

1 MOBILE TERMINALS

Everyday life without mobile terminals is almost unthinkable. With the increasing use of them for a multiplicity of tasks the equipment variants increase likewise. At present the most popular mobile terminals with the highest penetrations rates are mobile phones and Feature Phones. According to Kuhn (Kuhn, 2004) a mobile terminal is a portable computer-assisted communication system with independent data processing capacity, which can communicate as endpoint of a wireless connection with other IT-systems. “Independent data processing capacity” means that the mobile terminal is equipped with random access memory (RAM), a central processing unit (CPU) (this implies programmability), own network-independent power supply (battery) and the possibility to directly interact with human users by providing means for data input (e.g. keys, microphone, touchscreen) as well data output (e.g. display, light emitting diodes, speaker). Thus radio frequency identity tags (RFID), Smartcard etc. are not mobile terminals in the sense of the above definition because they don’t dispose over an independent power supply and a user interface. Computers integrated into machines (e.g. vending machines, machines in factories) are also not considered as mobile terminals because a human user can’t carry them along.

Mobile terminals differ particularly with regard to their size, their use and their performance for the execution of mobile applications. One reason for this is that mobile computing is a rather new technology so a market consolidation like that for personal desktop computers didn’t take place yet. There are also many application scenarios for mobile computers with different requirements so it is necessary to have mobile terminals especially developed for individual fields of applications. To get an impression of the great variety of mobile terminals one can refer to the WURFL\(^1\) project which assembles a profile database with information about capabilities of different mobile terminals: the file comprehends over 5,000 different profiles. Some software products for mobile terminals are available in customized versions for several hundred types of mobile terminals, e.g. “Yahoo!Go”\(^2\) a mobile applications that provides access to services like e-mail, news, weather forecasts and maps.

Nevertheless we can identify distinct classes of mobile terminals with common characteristics.

2 CLASSES OF TERMINALS

Within the range of the mobile terminals there are a number of different types. They differ in the characteristics listed above. There are different approaches to divide the terminals into different classes (see for example (Noesekabel and Lehner, 2002), (Meier, 2002) or (Roth, 2005)). The borders between the

\(^1\)http://wurfl.sourceforge.net/
\(^2\)http://mobile.yahoo.com/go
different terminal classes are seen differently by different authors and are often not clear. For example according to Brome and Zeman (Brome and Zeman, 2008) and Canalys (Canalys.com, 2004) the classification into Feature Phone and Smartphone is determined by whether additional software is executed directly by the operating system or by a runtime environment like Java ME. On the one hand this distinction isn’t obvious, on the other hand there are terminals which can do both.

Mobile terminals can differ in the following criteria (list is not complete):

- Size and weight
- Input modes
- Output modes
- Performance
- Kind of usage
- Communication capabilities
- Type of operating system
- Expandability

Some of these characteristics can be used to distinguish classes of terminals from each other. The following section gives criteria for the classification of mobile terminals into distinct classes without the need to do a deepened technical analysis. All classes regarded in our classification (see also Figure 1) have the ability for bidirectional wireless communication and have an independent power in the form of an accumulator. Hereby the often vague defined term Smartphone will be clearly separated from Feature Phone and Handheld.

2.1 Mobile Standard PC

The wireless mobile standard computers include laptops/notebooks, subnotebooks (also called netbooks) and Tablet-PC (including ultra mobile PC (UMPC)). Subnotebook is the name for a particularly small and light notebook, whose equipment is very similar to those in a full notebook, however has no optical drives (CD/DVD-ROM-drive) and a smaller keyboard. An UMPC is a Tablet-PC which is maximally 8 inches large, weighs maximally 900 grams and has a display with a resolution of at least 800x480 pixels; Keyboards can be attached externally. On these terminals the conventional desktop operating systems (Linux, Windows) can be used as well as the compatible software packages like office suites. Substantial distinguishing features are the size of chassis and display, the weight, the
integrated peripherals (e.g. optical Drive included) and the operation kind of the terminals.

Notebooks with wireless communication capabilities are mobile terminals according to definition given above. However in the following considerations we won’t focus on them because our article is more concerned with smaller and lighter terminals that can be used by a walking or standing user. Notebooks usually need a „stand space“ like a table or someone’s lap. Tablet PCs on the other hand are a special subclass of the mobile standard computers, because they have a touch-sensitive display and therefore can be used while standing with only one hand so they can be operated without a stand space.

2.2 Mobile Internet Devices

Beside the mobile standard PC there is a terminal class “Mobile Internet Devices” (MID) (also called Web-Tablet or mobile Thin Client) with similar equipment and display size, however clearly reduced function range and normally no mechanical drives. Characteristically these terminals are equipped like Handhelds (see below) with an operating system in the read-only memory (ROM) and are immediately ready for use after switching them on. Their function range is similar to the Handhelds and they are mainly used for browsing the web and as client for terminal server sessions. They depend mostly on a constant wireless network connection for the fulfilment of their tasks. The Web-Tablets (like the Tablet PC) can be operated through a touch-sensitive display, the mobile Thin Clients usually through a keyboard.

2.3 Handhelds

Handhelds are small computers, which — like the name already suggest — can be held in one hand. They are called often also Personal Digital Assistant (PDA). Historically these were originated from the pure organizers, which could be predominantly used for personal information management (PIM) applications like calendar, address book or personal notes. Today’s Handhelds have clearly higher performance and are usually equipped with multimedia capabilities. As written above, only Handhelds who are able to establish a wireless connection are regarded here. They are equipped usually with an operating system in the ROM and are immediately ready for use after switching on (no boot-time needed). Apart from the pre-installed programs in the ROM, users can install further applications in the RAM or the flash memory and attach various extension modules. Handhelds usually have a touch-sensitive display and can be used with a pen (or the fingers), or they have a text keyboard and navigation keys for it. Increasingly there are terminals which can be operated through both of them. Terminals are only classified as Handheld if they cannot communicate in communication networks for mobile telephony like GSM or UMTS.

2.4 Smartphones

The term “Smartphone” designates a combination of Feature Phone (see below) and Handheld with usually somewhat smaller dimensions, smaller display and partly reduced function range in comparison to the Handheld. They form their own class. On the one hand mobile computers which have the characteristics of Handhelds and the ability to communication over mobile telephony networks and have almost everywhere and „seamless“3 connection to the Internet belong to this class. On the other hand portable radio terminals with the characteristics of Feature Phones that have extended input modes belong to this class, too. The extended input mode can be provided by a touch-sensitive display (which can be used with the fingers or a pen) or a complete text keyboard.

2.5 Feature Phones

Feature Phones are mobile phones (see below) with usually somewhat larger display and extended function range. The primary use purpose of the Feature Phones lies in mobile communication, which includes both the voice communication and text-based communication (SMS, MMS, e-mail etc.). For this terminal class the ability for wireless data communication over the mobile communications networks (e.g. GPRS or UMTS) and the absence of the extended input modes are characteristic. These are limited to a number keyboard and a few additional keys. Particularly terminals which support not only the GSM standard but also UMTS are equipped with a Web browser and software to play back multimedia contents. On these terminals the user can install additional software. Since many terminals have proprietary operating systems, in most cases additional software is offered on basis of Java Micro

3 Seamless means that these terminals change unnoticed and without further user interaction independently from one radio cell to another (soft handover) and give the user the impression of only one large radio cell.
Edition (Java ME), which is supported by nearly every Feature Phone manufactured today.

2.6 Simple Phones

The classical mobile phones (also termed “cellular phone”) were developed primarily for mere voice communication. However nowadays text-based communication in the form of the short message service (SMS) is a standard feature of mobile phones. For Simple Phones it is characteristic that they are not able to use wireless data communication based on TCP/IP (e.g. browsing the internet or receive e-mail). The resolution and colour depth provided by displays of these terminals was significantly improved in the recent years. The displays became also bigger but they are usually noticeably smaller than those of Feature Phones. The function range of these terminals can only be extended with software in a small scale, e.g. by additional ringtones, pictures and multimedia animations or games, whereby these are strongly limited by the memory free space and the performance of the processor.

2.7 Special Terminals

Beyond the versatile usable standard terminals there are a number of specialized mobile terminals which cannot be arranged into the above classification. These are for example e-book readers for storage and reproduction of electronic books, electronic light pens, which can read bar code or printed texts and process electronically or bar code readers with number keyboard for the stocktaking in supermarkets: A worker scans the product code, enters the number of available products in shelf and transfers all immediately wirelessly to a server.

Navigation terminals for travellers became quite popular in the last few years. Those terminals can be used for locating as well as for route guidance. These terminals can calculate their own position based on the signals received from several satellites (Küpper, 2005). The global positioning system (GPS) of the US military or the Russian system GLONASS are already in use for many years and can be used nearly world-wide without fees. The projected Galileo system by the European Union is a civilian system that will provide some service for navigation free of charge; the full operation of this system is now scheduled to start in the year 2013. GPS (or others) receivers can be integrated into a mobile terminal but there are also separate GPS modules (so called “GPS mouse”) that can be connected to a mobile terminal using a cable or blue-tooth.

3 CONCLUSIONS

In this article we distinguished the mobile terminals clearly from other mobile devices. With the introduction of the term Feature Phone, which is used rarely by other authors, mobile phones can be divided selectively into different classes. The term Smartphone, which was until then fuzzy, become consequently clear. With the introduction of the class of special terminals, which comprehends all terminals for special tasks that cannot and should not be assigned to the standard classes, the classification of mobile terminals is altogether clear and complete.

ACKNOWLEDGEMENTS

This work has been funded by the Federal Ministry of Economics and Technology, Germany (BMWi, Contract No. 01MD06012). The responsibility for the content of this article lies solely with the authors.

REFERENCES

Brome, R., Zeman, E., 2008. Feature Phone definition. Phone Factor, LLC.
Kuhn, J., 2004. Commercial use of mobile applications: Results of the Delphi study “Mobile Business” (in German), Regensburg, Germany, University Regensburg.
Meier, R., 2002. Generation of customer value by mobile services: Potentials by communication and networking (in German), Wiesbaden, Germany, Deutscher Universitäts-Verlag.
IMPLEMENTING TRADING AGENTS FOR ADAPTABLE AND EVOLUTIVE UI-COTS COMPONENTS ARCHITECTURES

José Andrés Asensio, Luis Iribarne, Nicolás Padilla
*Applied Computing Group, University of Almería, Spain*
{jacortes, luis.iribarne, npadilla}@ual.es

Rosa Ayala
*Computers and Environmental Group, University of Almería, Spain*
rmayala@ual.es

Keywords: Trading agents, COTS User Interfaces, Model Transformation, Model-Driven Engineering, Cooperative Systems, e-Business Systems Modelling.

Abstract: Most of the complex (e-Business) information systems need to accomplish with the use of open standards. Environmental Management Systems (EMS), for instance, state the international regulations of the ISO 14000 family, which establish the requirements to be fulfilled by an EMS in order to be accepted as such. One of these requirements concerns the User Interfaces Development. Because of the variety of final users that interact in this sort of complex information system (politicians, technicians, administrators, and so on) and due to a great deal of information (some critical and confidential), it is important to have real and practical scientific/technical proposals in order to build fast and efficient information exploitation systems. The human-computer interaction (HCI) of these systems need user interfaces that adapt to the users profiles' habits, and with intelligent software agents that mediate by the users in the search processes, exploitation and decision-making tasks. In this work we present a part of the SOLERES-HCI, a framework of the Soleres Project for developing COTS user interfaces by using trading agents. Our studies are being applied for developing advanced EMS and approaching Model-driven engineering techniques to the UI-COTS development.

1 INTRODUCTION

Organizations and administrations that work with environmental information for specific actions such as territory management or planning and organization of natural resources (among others), need to have innovative quality information systems in order to guarantee the success of their everyday activities. Therefore, for the experts who operate the system not only it is important to have reliable and updated information that helps to make the most appropriate decisions, as it happens in critical complex systems. Due to the variety of final users that cooperate with each other and interact with the system for decision making (for instance, politicians, technicians, administrators, etc.), it is also important to have: (a) exploitation information systems (environmental, in this case) to facilitate the human-human and human-computer interaction and coordination; (b) intelligent user interfaces that adapt to the users profiles' habits and; (c) intelligent software agents that intercede on behalf of the users and facilitate the information interpretation tasks, the decision-making tasks and prediction/prevention tasks (which are the most important).

In this work, we present a part of the framework Soleres-HCI that supports all the human-computer interaction issues of a complex environmental management system (EMS). This portion of the framework concerns the developing of adaptable and evolutive user interfaces of the system by using: (a) trading agents that intercede between user agents and the information, (b) developing of traders that follow model-driven engineering perspectives, and (c) specialized Commercial Off-The Self (COTS) components for real-time user interface architectures.

The rest of the paper is structured as follows. Section 2 describes some issues for modelling user interfaces from UI-COTS. Section 3 continues with our SOLERES-HCI proposal, a human-computer interaction perspective based on trading (ISO/IEC, 1997) and software agents. Finally, Section 4 explains some of our ongoing research in the SOLERES project.
2 MODELLING UI-COTS

In recent years, different approaches for the design of user interfaces have been presented, most of them following model-driven UI designs, for instance:
(a) IDEAS (Interface Development Environment within OASIS) (Lozano et al., 2000): A methodology of UI development based on UML models.
(b) OVID (Object, View and Interaction Design) (Roberts et al., 1998): A methodology for UI design directed towards objects developed by IBM.
(c) TERESA (Paterno, 1999): A tool for the UI generation by using ConcurTaskTrees (CTT).
(d) WISDOM (Whitewater Interactive System Development with Object Models) (Nunes, 1998): A methodological proposal for UML-based UI.
(e) UMLi (Pinheiro, 2002): An extension of UML notation for the UI design.

In the SOLERES team there are also current works on UI modelling that use and extend the UML diagrams (Almendros and Iribarne, 2005) (Almendros and Iribarne, 2007). None of these works deal in depth with trading UI-COTS modelling.

On the other hand, our research aims to study UI following the approach of Component-based Software Development (CBSD) specialized for COTS components (Commercial Off-the-Shelf), (Meyers and Oberndorf, 2001). There are really few works showing realistic cases of IS development following the COTS paradigm or using multi-component UI-COTS. In (Iribarne et al., 2004) we developed an experiment COTS composition in Geographical Information Systems (GIS). However this approach doesn’t solve a dynamic and evolutive UI-COTS perspective.

Furthermore, the tendency during the last few years in CBSD is to facilitate the automatic integration of commercial components by means of the composition (assembly) of their parts. The advances in trading services (or traders) have played an important role for it. Inspired on the trading model for COTS components in open-distributed systems (Iribarne et al., 2005), (Iribarne et al., 2007) we develop a trading agent service for UI-COTS in EMS systems.

3 TRADING IN SOLERES-HCI

Soleres HCI is the framework of the SOLERES environmental management system, specialized in the human-computer interaction. This level of information system follows the paradigm of Computer Supported Cooperative Work (CSCW) and uses technology of agents and multiagent architectures.

For the process of data exploitation we identify and structure the type of queries and the sort of information suitable to be consulted, by using techniques of hierarchical decomposition (i.e., trees, cut and pruning) and neural networks.

Figure 1 shows the (short) architecture of the SOLERES information system in our project framework. At the user layer side (top), the system is designed in order to be used for environmental decision-making tasks and in cooperation among different people (system’s users) organized following different organized models (i.e. depending on their hierarchy).

This human-computer interaction and human-human interaction is guided by a cooperative system for the decision-taking tasks in group supported by a multi-agent architecture (next layer). Each user of the cooperative system has an UI agent whose function is
to adapt the UI to his/her needs by identifying his/her interaction habits with the UI. This UI agent mediates between the user and (a) the rest of the system’s users (who have their own UI agent), (b) the search information system (next layer) directly, or (c) an environmental software agent (environmental agent - MA).

A virtual consultant cooperates with other agents within a pre-established multi-agent architecture and respects a model of organization and cooperation (that is or project’s aim). The purpose of this cooperation (MA) is to facilitate the tasks of information exploitation: they interact with the search information system (we call an Environmental Process Unit - EPU) and filter the irrelevant information. The use of this sort of systems is quite useful because of its natural capacity to represent and implement organizational and social aspects that can help to identify and/or solve problems. Figure 2 shows internal details of an EPU.

For implementing the trading agent, we follow a Model-Driven Engineering (MDE) perspective based on the classical Model-driven Architecture (MDA) of the OMG. In this perspective, the trader model stays the three stages of the MDA: CIM/PIM/PSM (see Figure 3). A trader meta-model (MM) is defined for each stage, which describe the way to generate a diagrammatical model (UML) of the trader.

To translate a model into views (stages), we use model-transformation techniques. In our case, we use ATL for implementing the transformations of the trader. The language ATL (ATLAS transformation language) (Jouault and Kurtev, 2006) provides declarative and imperative constructs. The declarative part of ATL is based on rules. Such rules consist of a source pattern matched over source models and of a target pattern that creates target models for each match. Figure 4 shows a piece of the ATL code for the transformation CIM/PIM of one of the five interfaces of a trader agent: the Lookup interface. In the code, we can observe how associations, properties or operations of the Lookup class are created.

To implement the trading agent, we follow a Model-Driven Engineering (MDE) perspective based on the classical Model-driven Architecture (MDA) of the OMG. This perspective, the trader model stays the three stages of the MDA: CIM/PIM/PSM (see Figure 3). A trader meta-model (MM) is defined for each stage, which describe the way to generate a diagrammatical model (UML) of the trader.

To translate a model into views (stages), we use model-transformation techniques. In our case, we use ATL for implementing the transformations of the trader. The language ATL (ATLAS transformation language) (Jouault and Kurtev, 2006) provides declarative and imperative constructs. The declarative part of ATL is based on rules. Such rules consist of a source pattern matched over source models and of a target pattern that creates target models for each match. Figure 4 shows a piece of the ATL code for the transformation CIM/PIM of one of the five interfaces of a trader agent: the Lookup interface. In the code, we can observe how associations, properties or operations of the Lookup class are created.

```
rule Lookup {
  from #01 in_MM!Lookup
  using {
    O Lookup: out_MM!Operation = 'null';
    O get_lookup_if: out_MM!Operation = 'null';
    ...
  }
  to #01.out_MM!Class( name< 'Lookup', visibility<#public )
  do {
    O get_lookup_if< thisModule.createOperation('get_lookup_if',#public,'Lookup',false,false,1,1);
    O register_if< thisModule.createOperation('register_if',#public,'Register',false,false,1,1);
    A trader.C Lookup.C Trader< thisModule.createAssociation('trader.C Lookup.C Trader',#public,false);
    P trader.C Trader< thisModule.createProperty('trader',#private,'Trader'.false,false,1,1,false,'',false,false);
    ...
  }
}
```

Figure 4: A piece of the ATL transformation

Figure 5 shows the PIM trader model: a partial class diagram. In this framework we have also used the GMP/Eclipse to develop a tool for documenting CIM trader models. We use the OCL language of UML/OMG for describing the semantical restrictions.
of a trader. At the end, Java code generation from PSM trader models can be obtained by using a parser written in MOFScript.

For the implementation of the described framework we have used the Eclipse platform and JADE (http://jade.tilab.com/) for agents.

For space reasons, we have included here a part of the Soleres-HCI framework. A more complete version of the SOLERES project and details about the model transformations and implementations of the trading agent (p.e., Java code) are available at http://www.ual.es/acg/soleres.

4 ONGOING RESEARCH

The following are the specific objectives of our ongoing line of research.

Firstly, we plan to study the design of intelligent user interfaces. We will look into how to adapt the UI to different user’s profiles.

Our scientific objective (complement to previous one) is to study algorithms of dynamic-services composition in UI-COTS components architectures. We try to analyze this kind of component and its marketplace (i.e., how many exist, which types, how to define them, what kind of existing repositories, etc.).

Furthermore, starting from previous works (Iribarne et al., 2004), we plan to implement trading services of UI-COTS components and their applications in Cooperative Systems (specifically in environmental management systems).

ACKNOWLEDGEMENTS

This work has been partially supported by the EU (FEDER) and the Spanish MEC under grant of the project I+D TIN2007-61497 (SOLERES. A Spatio-Temporal Environmental Management System based on Neural-Networks, Agents and Soft. Components).

REFERENCES


BUSINESS AND
SOCIAL APPLICATIONS
DEALING WITH BUSINESS PROCESS EVOLUTION USING VERSIONS

Mohamed Amine Chaâbane, Eric Andonoff
Laboratoire IRIT, Université Toulouse 1, 2 rue du Doyen Gabriel Marty, 31042 Toulouse Cédex, France
MA.Chaabane@fsegs.rnu.tn, Andonoff@univ-tlse1.fr

Lotfi Bouzgenda, Rafik Bouaziz
Laboratoire MIRACL, ISIMS, Route de Tunis, km 10, BP 242, 3021 Sakeit Ezzit, Sfax, Tunisia
Lotfi.Bouzguenda@isimsf.rnu.tn, Raf.Bouaziz@fsegs.rnu.tn

Keywords: Version, Business Process Evolution, Meta-model, Petri net with Objects.

Abstract: Competition in which enterprises and organizations are involved nowadays imposes them to often make evolve their business processes in order to meet, as quickly as possible, new business or production requirements. This paper proposes to adopt a version-based approach to support these dynamic changes of business processes. This approach permits to keep chronological business process changes: it is then possible to allow several instances of a same business process to own different schemas, each one representing a possible schema for the considered business process. Consequently, this approach is very suitable to deal with long-term business process evolution: it does not necessarily impose the adaptation and migration of running instances of business processes to a new business process schema. The paper contribution is threefold. First, it defines a meta-model for designing versions of business processes considering the three dimensions of business processes: the informational, organizational and process dimensions. Then, it introduces a taxonomy of operations for business process version management. Finally, it proposes to formalize and visualize modeled versions of business processes using a Petri net-based formalism, namely Petri net with Objects.

1 INTRODUCTION

Nowadays, the importance of business processes in enterprises’ and organizations’ Information Systems (IS) is widely recognized. As a consequence, these last few years, there has been a shift from data-aware IS to process-aware IS (Aalst and al, 2007). However, even if important advances have been done in business process management, several problems are still to be dealt. Among them, the business process evolution problem that can be posed as follows: how to support dynamic change of business processes (Smith and Fingar, 2003), (Aalst and al, 2003-a)?

The competitive and dynamic worldwide economic context, in which enterprises and organizations are involved, lead them to often change and adapt their business processes in order to meet new business or production requirements. Consequently, the business process evolution problem is really a relevant problem. This problem has mainly been addressed in the Workflow context using two main approaches: an adaptive-based approach and a version-based approach. The adaptive-based approach consists in defining a set of operations supporting both workflow process schema changes, and adaptation and migration of their corresponding instances (Casatia and al, 1996), (Reichert and Dadam, 1998), (Kammer and al, 2000), (Rinderle and al, 2004). In this approach, only one schema is kept for all modelled workflow processes. This approach has been investigated intensively and its implementations, ADEPT (Reichert and Dadam, 1998) and JOpera (Heinis and al, 2005), are probably the most successful Workflow Management Systems (WFMS) regarding workflow process’ schema evolution.

In the version-based approach, different instances of a same workflow process can have different schemas. Thus, it is possible to distinguish between temporary and permanent updates for workflow processes since it is possible to keep track of chronological workflow process changes, each
one representing a possible schema for the considered workflow process.

In the workflow context, where long-term processes are involved, adaptation and migration of workflow process instances according to a new schema are not always easy and are sometimes impossible (Casati and al, 1996). So, it is important to be able to manage different schemas for a workflow process in order to allow several instances of this workflow process to own different schemas (Kradolfer and Geppert, 1999). Thus, the version-based approach is a promising solution to deal with business process evolution.

Versions are used in several fields of computer science in which was highlighted the need to describe evolution of real world entities over time. Thus, versions are used in the database field mainly in object-oriented databases (Cellary and Jomier, 1990), (Sciorc, 1994), or scientific databases (Chen and al, 1996) but also for specific database application fields such as computer aided design or computer aided manufacturing (Chou and Kim, 1986), (Katz, 1990). Versions are also used in software engineering to handle software configurations (Kimball and Larson, 1991). Versions are also considered in conceptual models such as the Entity Relationship model (Roddick and al, 1993) or the OMT model (Andonoff and al, 1996).

Although versions are used in several areas of computer science, to the best of our knowledge, only few efforts have been put on version management in the business process (workflow) context (in the remainder of the paper, the terms workflow and business process will be used equally).

We distinguish two main contributions about versions of business processes in literature. (Kradolfer and Geppert, 1999) have proposed to deal with dynamic workflow evolution, i.e. modification of workflow process schemas in the presence of active workflow process instances, introducing versions of workflow process schemas. This work has defined a set of operations for workflow process schema modification and, if possible, a strategy for migration of workflow process instances. Recently, (Zhao and Liu, 2007) have also defended the advantages of a version-based approach to face business process evolution. More precisely, this work proposes to model versions of workflow process schemas using graphs. It also presents a set of operations enabling updates of graphs and defines two strategies to extract versions of workflow process schemas from these graphs.

We believe that these two propositions need to be revisited. Indeed, both (Kradolfer and Geppert, 1999) and (Zhao and Liu, 2007) addressed the issue of business process versioning only considering, what is called in the workflow literature, “process model”. Such a model describes tasks involved in the process and their coordination. But, using only this model is not enough to have a comprehensive description of business processes (Aalst, 1999). Two others models have to be considered: the organizational and the informational models. The organizational model structures the business process actors and authorizes them, through the notion of role, to perform tasks making up the process. The informational model defines the structure of the documents and data required and produced by the process. These two models are glued together with the process model since, in addition to the tasks and their coordination, the process model also defines the required resources (information, actors) to perform the tasks.

Consequently, this paper proposes to revisit the business process evolution problem using a version-based approach and considering both organizational, informational and process models of business processes. More precisely, this paper introduces:

- A meta-model for designing versions of business processes;
- A taxonomy of operations for business process version management;
- A formalization and a visualization of versions of business processes designed with the previous meta-model.

The remainder of this paper is organized as follows. Section 2 introduces the Business Process (BP) meta-model we use for designing business process, while section 3 introduces the Versioned Business Process (VBP) meta-model we propose for business process versioning. More precisely, section 3 first recalls the notion of version, then presents the versioning kit we propose for handling versions of business processes, and finally explains how the kit is merged with the BP meta-model to define the VBP meta-model. This section also gives an example of business process versioning. Section 4 is dedicated to the dynamic aspects of the meta-model: it presents a taxonomy of operations for business process version management. Section 5 presents our proposition for both formalization and visualization of workflow process versions using a formal model, namely Petri Net with Objects (Sibertin, 1985). Finally, section 6 stands our contribution according to related works and then concludes the paper.
2 MODELING BUSINESS PROCESSES

As mentioned before, a business process meta-model must allow the expression of three complementary aspects, usually described through three different interacting models: the organizational, informational and process models. The main important model is the process model which defines component tasks and their coordination, but this model also refers to the organizational model and to the informational model defining required and produced resources before and after tasks execution (Aalst, 1999).

Another important requirement for such a meta-model is its simplicity and efficiency: it must be comprehensive and must define the core (basic) concepts of the three complementary aspects of business processes: it must play the role of a Business Process Virtual Meta-model, i.e. a minimal meta-model for the design of business processes. This idea of Business Process Virtual Meta-model is the same as the one of Workflow Virtual Machine introduced in (Fernandes and al, 2004) to deal with the development of a Workflow Management System (WiMS) that supports changes in its workflow definition language(s).

But does such a meta-model for business process modelling (i.e. meeting the previous requirements) already exist, or do we have to define a new one by ourselves?

Despite the standardization efforts of the Workflow Management Coalition (WiMC), different workflow or business meta-models exist in literature. The used vocabulary differs from one model to another, and yet, so far, the workflow and business process community seem to not have reached an agreement on which model to adopt, even if XPDL, BPMN and BPEL are standards recommended by the WiMC.

Some process and workflow meta-models proposed in literature mainly focus on the process model -i.e. tasks description and their coordination- (e.g. (Zhao and Liu, 2007), BPEL, XPDL). Others also consider the informational model in addition to the process model (e.g. (Casati and al, 1995), (Kradolfer and Geppert, 1999), (Vossen and Weske, 1999), (Aalst and al, 2004)). Finally, some meta-models have a comprehensive approach for business process modeling considering the three complementary aspects (e.g. FlowMark and its successors MQSeries Workflow and WebSphere MQ Workflow (Leymann and Roller, 1999), Exotica (Mohan and al, 1995), OpenFlow (Halliday and al, 2001)). For instance, as illustrated in (Rosemann and zur Muehlen, 1998), the FlowMark meta-model proposes a very detailed description of workflow processes along with involved data flows and actors. However, these meta-models are very complex, specially with respect to the organizational dimension.

Consequently, we have defined our own meta-model which fulfills the previous requirements: (i) a comprehensive meta-model considering three complementary aspects of business processes (organizational, informational and process models), and (ii) a business process virtual meta-model as it defines the core (basic) concepts of the three complementary aspects of business processes. This meta-model is shown in the UML diagram of figure 1.

In this UML meta-model, a business process is either a composite process or an atomic process. A composite process is itself recursively composed of atomic or composite processes. It also uses a control pattern, which participates to the definition of business process coordination. In our meta-model, and as in, for instance (Manolescu, 2001), the main control patterns described in the literature are provided. Some of them are conditional (e.g. if, while…), while others are not (e.g. sequence, fork…). Their semantics is the following:

- **Sequence** pattern: it permits the execution of processes in a sequential order;
- **If** pattern: it allows processes execution according to a condition;
- **Fork** pattern: it spawns the parallel execution of processes and waits for the first to finish;
- **Join** pattern: it spawns the parallel execution of processes but waits for all of them before completing;
- **While** and **Repeat** patterns: they cyclically execute a process while or until a condition is achieved.

![Figure 1: The Business Process meta-model.](image)

An atomic process corresponds to a task to perform. It can have pre-conditions and post-conditions, and executes one or several actions. An atomic process is performed by a role (belonging to
the organizational model) and consumes and/or produces informational resources (belonging to the informational model). Informational resources correspond to system data, process data (i.e. data, document or form), and application data (i.e. database and data repository). A role is played by an actor belonging to some organizational units. An actor is a human resource or not (machine or software). Finally, an actor may be internal or external.

Going back to control patterns, our meta-model only includes low level (basic) control patterns; all the high level workflow patterns of (Aalst and al, 2003-b) are not considered here (they are much more complex than what we need). In this way, the meta-model we propose could be seen as a Business Process Virtual Meta-model gathering the core (basic) concepts of business process models.

3 MODELING VERSIONS OF BUSINESS PROCESSES

First, this section briefly recalls the version notion as it is introduced in object-oriented databases and software engineering. Then, this section presents the Versioned Business Process (VBP) meta-model: it consists of a versioning kit to handle versions of business processes which is merged with the BP meta-model introduced before. Finally, this section illustrates the VBP meta-model instantiation to design versions of business processes.

3.1 Concept of Version

A real world entity has characteristics that may evolve during its life cycle: it has different successive states. In object-oriented database systems that provide version management, this entity is described by a set of objects called versions. A version corresponds to one of the significant entity states. Then, it is possible to manage several entity states (neither only the last one as in classical databases nor all the states as in temporal databases).

As illustrated in figure 2, the entity versions are linked by a derivation link; they form a version derivation hierarchy. When created, an entity is described by only one version. The definition of every new entity version is done by derivation from a previous one. Such versions are called derived versions (e.g. E1.v1 is a derived version from E1.v0). Several versions may be derived from the same previous one. They are called alternatives (e.g. E1.v2 and E1.v3 are alternatives derived from E1.v1).

A version is either frozen or working. A frozen version describes a significant and final state of an entity. A frozen version may be deleted but not updated. To describe a new state of this entity, we have to derive a new version (from the frozen one). A working version is a version that temporarily describes one of the entity states. It may be deleted and updated to describe a next entity state. The previous state is lost to the benefit of the next one.

3.2 The Versioned Business Process Meta-model

This meta-model consists of a versioning kit to handle versions of business processes, which is merged to the BP meta-model previously introduced.

3.2.1 Versioning Kit

This kit is very simple: it is composed of a class and a set of properties and relationships that make classes of the previous meta-model “versionable”. A “versionable” class is a class whose instances are versions (Katz, 1990).

Thus, for each of these “versionable” classes, we define a new class which contains versions, called “Version of…”. We also specify two new relationships: (i) the is_version_of relationship which links a class to its corresponding “Version of…” class, and (ii) the derived_from relationship which describes version derivation hierarchies. This latter relationship is reflexive. The underlying idea of our proposition is to describe both entities and their corresponding versions as indicated in figure 2. Consequently, (i) versions are therefore involved in the process definition, and (ii) a couple (version, entity) is obviously created when the first version of an entity is created. Regarding properties of these “Version of…” classes, we introduce the classical version number, creator name, creation date and status properties (Sciore, 1994).

3.2.2 Merging the Versioning Kit with the Business Process Meta-model

Regarding the process model, we propose to keep versions for only two classes: the Atomic Process
and the Business Process classes. It is indeed interesting to keep changes history for both atomic processes (i.e. tasks) and workflow processes since these changes correspond to changes in the way that business is carried out. At the atomic (task) level, versions describe evolutions in activity realization while at the business process level, versions describe evolutions in work organization (i.e. coordination of activities). We defend the idea that atomic process (task) and business processes versioning is enough to help organizations to face the fast changing environment in which they are involved nowadays.

Regarding the other models, it is necessary to handle versions for the Informational resource class from the informational model, and versions for the Role class from the organizational model. Regarding this latter model, it is also possible to handle versions for the Actor and Organizational Unit classes. However, keeping changes history for these two classes is, in our opinion, quite useless to handle versions of business processes.

Figure 3 below presents the new obtained meta-model in terms of classes and relationships.

![Figure 3: The Versioned Business Process meta-model.](image)

### 3.3 Example

In order to illustrate the VBP meta-model instantiation, we propose to use the example introduced by (Zhao and Liu, 2007). Because of space limitation, we only focus on the instantiation of the process model of this example.

This example describes a production business process and involves a factory, which owns one production pipeline following the business process shown in figure 4(a). It includes several activities: production scheduling, production using a work centre, quality checking and packaging. In order to increase its productivity, the factory decides to add a new work centre. The business process is then updated as shown in figure 4(b). If one of the two work centres, for instance work centre #1 (Pc#1), has a technical problem and consequently is removed from the process, two solutions are proposed to attempt keeping the production output: fixing unqualified products or using employees for manual production. The business process is then updated as shown in figure 4(c) and 4(d).

![Figure 4: Change in the Production BP.](image)

The solution we provide to model theses derivation hierarchies consists in instantiating the VBP meta-model. The Business Process, Atomic Process, Role and Informational resource ”versionable” classes and their “Version of...” corresponding classes are involved in this instantiation, along with the Composite Process and Control Pattern non “versionable” classes. This instantiation is visualized in figure 5.

![Figure 5: Instantiation of the VBP Meta-Model.](image)
4 OPERATIONS FOR BUSINESS PROCESS VERSIONING

In this section, we introduce a taxonomy of operations for business process versioning. These operations are defined as methods in the “Version of …” classes (“versionable” classes). They correspond to classical operations for versions (Katz, 1990): create, derive, delete, update and froze, but this taxonomy also includes operations for version selection. Of course, create, delete and update are also available for the other classes of the meta-model (non “versionable” classes), but their presentation is out of the scope of the paper.

This section introduces the create, derive, delete, update and froze operations first giving a state chart which indicates when these operations are available, and second detailing the actions they perform according to the classes in which they are defined. Moreover, this section also discusses about version selection, more precisely business process version selection.

4.1 State Chart for Versions

The UML state chart of figure 6 indicates when these operations are available. Some of them are available whatever the state of versions on which they are applied, while others are only available in some cases. In this state chart, each operation is described using the notation Operation:Event/Action whose meaning is “for Operation when Event is triggered then Action is performed”.

![State Chart for Versions](image)

When the create event is triggered by the a version designer, the to_create action is performed to both create the entity and its corresponding first version. The state of the created version is Working. In this state, the version can be updated (update event and to-update action).

It also can be deleted (delete event and to-delete action): its state is then the final state of the chart. It also can be frozen (freeze event and to-freeze action): its state is then Frozen. Triggering the freeze event, the designer means that the considered version is definitive and does not need additional updates. A frozen version (i.e. a version in a Frozen state) can be deleted or can serve as a basis for the creation of a new version using the derive event and to-derive operation. This new created version has the same value as the version from which it is derived from: its state is Working.

In addition to the previous state chart, these operations require further details. For instance, the Create and Update operations permit to add and delete references to the components of versions. These components change according to the considered type of versions: versions of business processes, versions of atomic processes (tasks), versions of informational resources or versions of roles. Regarding the Derive operation, it can trigger the derivation of versions of its components. The sections below gives additional details for these operations.

4.2 Creating and Updating Versions

Table 1 and 2 below give the semantics of these two operations (Create and Update) according to the classes in which they are defined. The four “Version of…” classes are considered.

<table>
<thead>
<tr>
<th>Business Process</th>
<th>Atomic Process</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Change structure</td>
<td>1. Change conditions</td>
</tr>
<tr>
<td>1.1. add/delete composite process in the workflow process structure</td>
<td>1.1. add/delete pre-conditions (has-pre-conditions relationship)</td>
</tr>
<tr>
<td>1.2. add/delete atomic process in the workflow process structure</td>
<td>1.2. add/delete post-conditions (has-post-conditions relationship)</td>
</tr>
<tr>
<td>2. Change pattern</td>
<td>2. Change action</td>
</tr>
<tr>
<td>2.1. choose a pattern for a composite process (use relationship)</td>
<td>2.1. add/delete actions (contains relationship)</td>
</tr>
<tr>
<td>3. Change information</td>
<td>3.1. add/delete input information (consumes relationship)</td>
</tr>
<tr>
<td>3.2. add/delete output information (produces relationship)</td>
<td>4. Change role</td>
</tr>
<tr>
<td>4.1. add/delete roles (references relationship)</td>
<td></td>
</tr>
</tbody>
</table>

These two tables indicate that Create and Update operations change according to the classes in which
they are defined. However, they share the same general idea that is to give values to properties and relationships of the considered classes. Moreover, relationships referencing versions may only reference frozen versions (i.e., versions in the Frozen state).

<table>
<thead>
<tr>
<th>Informational Resource</th>
<th>Role</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Change software</td>
<td>1. Change actors</td>
</tr>
<tr>
<td>1.1. add/delete software (uses relationship)</td>
<td>1.1. add/delete actors (played_by relationship)</td>
</tr>
<tr>
<td>2. Change the structure of information resource</td>
<td>2. Change organization</td>
</tr>
<tr>
<td>2.1. add/delete organizational units (belongs_to relationship)</td>
<td>2.1. add/delete organizational units (belongs_to relationship)</td>
</tr>
</tbody>
</table>

**4.3 Derivation of Versions**

The Derive operation allows the creation of a new version from an existing frozen one. The new created version is a working version (its state is working). Before being updated, the value of this new created version is the same as the derived one. Moreover, derivation of a version may trigger the derivation of other versions, which are linked to the derived one. Figure 7 below illustrates this derivation propagation.

**4.4 Selection of Versions**

In addition to the previous presented operations, we also propose specific operations for version selection and version hierarchy selection: Select, Slice, Display, among others. Because of space limitation, the paper only details the version selection operation and illustrates its use for business process version selection.

**4.4.1 Select Operation**

This operation allows the selection of versions. Its syntax is: \texttt{Select(Class,Predicate)} where Class is a name of a VBP class containing versions (i.e., a “versionable” class) and Predicate a condition permitting the filtering of versions.

The result of this operation is a set of versions verifying the predicate along with versions and/or objects that are (directly or not) linked to it by a relationship. In other words, the result of the Select operation is a set of instances of the VBP meta-model linked (directly or not) to a version belonging to the “versionable” class on which the Select operation is performed. We call such a group of instances VBP-instances. This notion of VBP-instances corresponds to the notion of Configuration introduced for handling versions in Software Engineering (Kimball and Larson, 1991). It is also close to the notion of Database Version introduced in (Cellary and Jomier, 1990) in order to reduce the complexity of version management in object-oriented databases.

**4.4.2 VBP-Trees for representing VBP-Instances**

Regarding business process version selection, the result of a Select operation performed to the Version of Business Process class is a set of business process versions verifying the predicate along with instances (versions and/or objects) of the Composite Process, Control Pattern, Version of Atomic Process, Version of Informational Resource and Version of Role classes which are (directly or not) linked to them.

In this case, a VBP-instance corresponds to a business process version along with versions and/or objects linked to it. It can be represented as what we call an VBP-Tree from which we distinguish two kinds of nodes: terminal nodes (leaves) and nonterminal nodes. Terminal nodes correspond to VBP atomic processes while non terminal nodes correspond to VBP composite processes. A non terminal node is described by the following data structure:

- **NodeName**: name of the node (corresponds to the name of the corresponding composite process);
- **CPName**: name of the control pattern used for the composite process;
- **Condition**: optional property associated to conditional control patterns;
- **SetOfNodes**: set of nodes (terminal or non terminal) composing it.

A terminal node is described by the following data structure:
- **NodeName**: name of the node (corresponds to the name of the corresponding atomic process);
- **SetOfActions**: set of actions to perform;
- **PreCondition**: condition associated to the execution of the atomic process that the node represents;
- **PostCondition**: condition associated to the atomic process after execution of actions of the node;
- **ConsumesInformation**: set of informational resources required to perform actions of the node;
- **ProduceInformation**: set of informational resources produced by the performing the actions of the node;
- **PlayedBy**: role defining a set of actors able to perform the actions of the node.

For instance, the VBP-Tree corresponding to the third version of the Production business process (vbp3 i.e. Production.v3) introduced in section 3.3 is visualized in figure 8.

This VBP-Tree illustrates the structure of the considered business process distinguishing terminal nodes (visualized as ellipses) from non terminal nodes (visualized as rectangles).

In fact, figure 8 only gives a simplified view of the VBP-Tree since nodes are not described in details (according their corresponding structures defined before).

![Figure 8: VBP-Tree for Production v3.](image)

The function implementing the mapping from a VBP-instance to a VBP-Tree uses the mapping rules given in Table 3 below.

<table>
<thead>
<tr>
<th>VBP meta-model concepts</th>
<th>VBP-Tree concepts</th>
</tr>
</thead>
<tbody>
<tr>
<td>Instance of Version of Business Process class</td>
<td>VBP-Tree</td>
</tr>
<tr>
<td>Instance of Composite Process class</td>
<td>Non Terminal node</td>
</tr>
<tr>
<td>Instance of Version of Atomic Process class</td>
<td>Terminal node</td>
</tr>
</tbody>
</table>

Moreover, this function uses a set of functions permitting the handling of processes and nodes:
- **IsAtomicProcess(i)** indicates if i is an instance of the Version of Atomic Process class;
- **BuildTerminalNode(i)** returns the corresponding terminal node of an atomic process i taking into account the relationships flowing from i (has_pre-conditions, ...);
- **BuildNonTerminalNode(i)** returns the corresponding non terminal node of the composite process i using the relationship flowing from i (uses);
- **AddNode(n, tr)** adds the node n to a VBP-Tree tr.

This function is the following.

Function BuildVBP-Tree (i:VBP-Instance):VBP-Tree
Local n:Node
Begin
  If IsAtomicProcess(i) Then
    n = BuildTerminalNode(i)
    BuildVBP-Tree = AddNode(n, tr)
  Else
    -- i is a composite process
    n = BuildNonTerminalNode(i)
    BuildVBP-Tree = AddNode(n, tr)
    For Each c ∈ IsComposedOf(i)
      BuildVBP-Tree = BuildVBP-Tree(c)
    Next c
  End If
End

5 FORMALIZING BUSINESS PROCESS VERSIONS: FROM VBP-TREE TO PNO

Representing versions of business processes as VBP-Tree is not sufficient to visualize and formalize the semantics of the modeled versions of business processes. To compensate this drawback, we propose to use a Petri net-based formalism, namely Petri Nets with Objects (PNO) (Sibertin, 1985) for
workflow process version visualization and formalization.

This section first presents the PNO formalism and gives the reasons of the choice of this language for workflow process versions. Then, this section explains the mapping from a VBP-Tree onto a PNO.

5.1 Petri-net with Objects

5.1.1 What are PNO?

Petri Nets with Objects (PNOs) (Sibertin, 1985) are a formalism combining coherently Petri nets (PN) technology and the Object-Oriented (OO) approach. While PN are very suitable to express the dynamic behavior of a system, the OO approach permits the modeling and the structuring of its active (actor) and passive (information) entities. In a conventional PN, tokens are atomic, whereas they are objects in a PNO. As any PN, a PNO is made up of places, arcs and transitions, but in a PNO, they are labeled with inscriptions referring to the handled objects. More precisely, a PNO features the following additional characteristics:

- Places are typed. The type of a place is a (list of) type of an (list of) object(s). A token is a value matching the type of a place such as a (list of) constant (e.g. 2 or ‘hello’), an instance of an object class, or a reference towards such an instance. The value of a place is a set of tokens it contains;
- Arcs are labeled with parameters. Each arc is labeled with a (list of) variable(s) of the same type, as the place the arc is connected to. The variables on the arcs surrounding a transition serve as formal parameters of that transition and define the flow of tokens from input to output places. Arcs from places to a transition determine the possible condition of the transition: a transition may occur (or is possible) if there exists a binding of its input variables with tokens lying in its input places;
- Each transition is a complex structure made up of three components: a precondition, one (or several) action(s) and emission rules. A transition may be guarded by a precondition, i.e. a side-effect free Boolean expression involving input variables. In this case, the transition is only permitted by a binding if this binding evaluates the precondition to be true. Passing through a transition depends on the precondition, on the location of tokens and also on their value. A transition also includes one or several actions, which consists of a piece of code in which transitions’ variables may appear and object methods may be invoked. These actions are executed at each occurrence of the transition and they process the values of tokens. Finally, a transition may include a set of emission rules i.e. side-effect free Boolean expressions that determine the output arcs that are actually activated after the execution of the action.

5.1.2 Motivations for using PNO

Petri nets are widely used for workflow specification (Aalst, 1998). Several good reasons justify their use:

- An appropriate expressive power that permits the description of the different tasks involved in a workflow process and their coordination;
- A graphical representation that eases the workflow process specification;
- An operational semantics making an easy mapping from specification to implementation possible;
- Theoretical foundations enabling analysis and validation of behavioral properties and simulation facilities.

Unfortunately, conventional Petri nets focus on the process definition and do not perfectly capture the organizational and the informational dimensions of business processes. As mentioned before, PNO extend Petri nets by integrating high-level data structures represented as objects, and, therefore provide the possibility to integrate in a coherent way the two missing dimensions. Thus, using PNO, actors/roles of the organizational model are directly represented as objects and they may be invoked through methods in the action part of a transition. In the same way, data and documents (from the informational model) are also represented by objects flowing in the PNOs and transformed by transitions.

Consequently, we use PNO as a graphical tool to visualize versions of business processes, and as a formal tool to define executable specifications in order to analyze, simulate, check and validate workflow process versions.

5.2 From VBP-Trees to PNOs

Table 4 and figure 9 give the mapping rules in order to obtain, from a VBP-Tree, i.e. a VBP-instance, the corresponding Petri net with objects. We distinguish mapping rules for concepts from mapping rules for control patterns. Table 4 introduces mapping rules for concepts while figure 9 presents mapping rules for control patterns.
### Table 4: Mapping Rules for Concepts.

<table>
<thead>
<tr>
<th>VBP-Tree concepts</th>
<th>PNO concepts</th>
</tr>
</thead>
<tbody>
<tr>
<td>Name of a Terminal node</td>
<td>Name of a transition</td>
</tr>
<tr>
<td>Set of Actions of a Terminal node</td>
<td>Actions of a transition</td>
</tr>
<tr>
<td>Pre Condition of a Terminal node</td>
<td>Pre-condition of a transition</td>
</tr>
<tr>
<td>Post Condition of a Terminal node</td>
<td>Emission rule of a transition</td>
</tr>
<tr>
<td>Consumes Information of a Terminal node</td>
<td>Begin place of a transition</td>
</tr>
<tr>
<td>Produce Information of a Terminal node</td>
<td>End place of a transition</td>
</tr>
<tr>
<td>Played By of a Terminal node</td>
<td>Begin place of a transition representing a role</td>
</tr>
</tbody>
</table>

Figure 9 below details how modeled control patterns are represented using PNOs. In this figure, P, P1 and P2 correspond to business processes while condition corresponds to a condition used in conditional control patterns. Finally, Empty is a transition for which no actions are executed.

![Diagram](image)

Figure 9: Mapping Rules for Control Patterns.

We also provide a function for building a PNO from a VBP-Tree. This function uses mapping rules presented in table 4 and figure 9 for defining transitions of the PNO and their coordination.

Moreover, this function uses a set of functions permitting the handling of a tree:

- ListOfChildren(n) returns the children of a node n (non-terminal or terminal nodes);
- ListOfLeaves(t) returns the terminal nodes (leaves) of a tree t;

and also a set of functions for building transitions and their coordination:

- BuildTransition returns the corresponding transition to a node using mapping rules defined in table 4;
- BuildPattern uses mapping rules defined in figure 9 to return the corresponding PNO according to the control pattern specified in a node;
- AddTransition (tr,PNO): add a transition tr to a PNO;
- AddPattern (pa,PNO): adds a pattern pa to a PNO.

This function is the following.

```plaintext
Function BuildPNO (n:Node):PNO
Local c:Node; tr:Transition
Global t:VBP-Tree
Begin
If n ∈ ListOfLeaves(t) Then
  tr = BuildTransition(n)
  BuildPNO = AddTransition(tr,BuildPNO)
Else
  -- n is a non terminal node
  pa = BuildPattern(n)
  BuildPNO = AddPattern(pa, BuildPNO)
For Each c ∈ ListOfChildren(n)
  BuildPNO = BuildPNO(c)
Next c
End If
End
```

### 6 CONCLUSIONS

As mentioned in the introduction, the problem stated as “how to support dynamic change of business process” has already been addressed in the workflow context. We distinguish two main approaches to deal with this problem.

Concerning the adaptive-based approach, relevant works in this area propose solutions to deal with workflow schemas changes, adaptation and migration of their corresponding instances. (Casatia and al, 1996) presents a workflow modification language that supports updates of workflow
DEALING WITH BUSINESS PROCESS EVOLUTION USING VERSIONS

schemas. It also defines a set of evolution policies that a workflow administrator can adopt to manage instances of updated workflow schemas in order to migrate (or not) them as instances of the new schema. Three mains policies are defined: abort, flush and progressive. (Kammer and al, 2000) investigates exception handling as a way to support dynamic change to workflow process schemas. Consequently, it introduces a taxonomy for exceptions and defines functionalities that Workflow Management Systems must have in order to be able to deal with these exceptions. The ADEPTflex project (Reichert and Dadam, 1998), (Rinderle and al, 2004) extensively studies process schema evolution. This work formally defines change operations for both process schemas and workflow instances as well as related migration policies in handling potential conflicts. We can also mention van der Aalst’s work to address dynamic change of workflow (Aalst, 2001). This work uses a generic process model to describe a family of variants of a same workflow process and the notion of inheritance is used to link these different variants. In the same vein, (Adams and al, 2006) proposes, for dealing with dynamic evolution in workflows, to use accepted ideas of how people actually work to define sets of worklets (i.e. processes) and a strategy for runtime selection of a specific worklet.

However, none of these works mention the notion of workflow versions. Consequently, none of them enables several different schemas of a same workflow process to conjointly exist.

Relevant works from the version-based approach allow to different instances of a same workflow process to own different schemas. Two main contributions are relevant from this approach. First, (Kradolfer and Geppert, 1999) have proposed to deal with dynamic workflow evolution, i.e. modification of workflow process schemas in the presence of active workflow process instances, introducing versions of workflow process schemas. This work has defined a set of operations for workflow process schema modification and a strategy for migration of workflow process instances. Second, and more recently, (Zhao and Liu, 2007) have also defended the advantages of a version-based approach to face business process evolution. More precisely, this work proposes to model versions of workflow process schemas using a graph. It also presents a set of operations enabling to update this graph and defines two strategies to extract versions of workflow process schemas from this graph.

However, these two works only consider the process model of workflow. They do not integrate the two other dimensions of workflow, that are the informational and the organizational dimensions.

Consequently, this paper revisits the dynamic change of business process issue following a version-based approach and considering the organizational, informational and process models of business processes. More precisely, it introduces:

- A meta-model for designing versions of business processes;
- A taxonomy of operations for business process version management;
- A formalization and a visualization, using Petri net with Objects, of versions of business processes, designed with the previous meta-model.

Our solution has the following advantages:

- It permits a comprehensive modeling of business processes considering the three dimensions of business processes;
- The VBP meta-model is simple: it only integrates core concepts for both business process modeling and business process versioning (our versioning kit is very simple),
- Dynamics aspects of business process version management are investigated in depth according to the state of the art for versions in databases;
- It provides rules and algorithms to derive modeled versions of business processes onto Petri net with objects specifications.

As future work, we have planned to implement the VBP meta-model in order to model version of business processes and to derive versions of business processes specified using BPEL.

REFERENCES


A THEORY-DRIVEN FRAMEWORK FOR CONSUMERS TO ADOPT M-COMMERCE DEVICES

Vincent Cho, Humphry Hung and Y. H. Wong
Department of Management and Marketing, The Hong Kong Polytechnic University, Hung Hom, Hong Kong
msvcho@polyu.edu.hk, mshung@polyu.edu.hk, msyhwong@polyu.edu.hk

Keywords: Mobile commerce.

Abstract: M-commerce (mobile commerce) is the buying and selling of goods and services through wireless handheld electronic devices such as portable personal computers, mobile phones and personal digital assistants. This paper proposes a theory-driven examination of the adoption of M-commerce devices (MCD) by consumers in their on-line purchase processes. By integrating the concepts of the options model with the major ideas of the technology acceptance model (TAM), we identify four M’s (merits, maturity, maneuverability and mentality) as the influencing factors of the adoption of MCD. Based on the generic attributes of m-commerce, we further identify two M’s, matching and mobility, as the antecedents of these influencing factors. We then propose a conceptual model of the adoption of MCD by consumers. Because of the ubiquitous nature of m-commerce, the proposed framework would have universal implications and would make significant contributions to a more in-depth understanding of the spread and acceptability of m-commerce.

1 INTRODUCTION

It is estimated that by 2007, the total number of the Internet users in the world will be over 1.4 billion and the percentage of wireless users is projected to take up about 57% of the vast number (Magura, 2003). Most people anticipate that the next-generation commerce will emerge from traditional commerce, to PC-based e-commerce, and eventually to mobile commerce (Chircu & Kauffman, 2000; Ellis-Chadwick et al., 2000, Miller, 2002). M-commerce (mobile commerce) is the buying and selling of goods and services through wireless handheld devices such as mobile phone and personal digital assistants. It is an extension, rather than a complete replacement, of PC-based e-commerce (electronic commerce) and allows users to interact with other users or businesses in a wireless mode, anytime and anywhere (Balasubramanian et al, 2002; Samuelsson & Dholakia, 2003). It is very likely that PC-based e-commerce will still prevail for a relatively long period of time in spite of the trend that more and more people will choose to adopt m-commerce for their purchases (Miller, 2002).

As content delivery over mobile devices becomes faster, more secure, and efficient, there is wide speculation that m-commerce will surpass PC-based e-commerce as the preferred method of choice for digital commerce transactions. The industries and services affected by m-commerce will include: 1) information services, which include the delivery of financial news, sports, horse racing and traffic updates to users; 2) financial services, which includes e-banking (when customers use their handheld devices to access their accounts and pay their bills) as well as brokerage and investment services, in which stock quotes can be displayed and trading conducted from mobile devices; 3) telecommunications, in which service charges, bill payment and account reviews can all be conducted from mobile devices; and 4) retail consumers are given the ability and opportunity to place orders and pay through mobile devices.

The emerging technologies behind m-commerce is based on the Wireless Application Protocol (WAP) and high speed wireless network such as 3G, 3.5G and 4G (Cowles, Kiecker, & Little, 2002; Watson et al., 2002). These technologies enable users to download video/audio information seamlessly. The focus of this research framework is on the consumers’ adoption of m-commerce devices (MCD), which are equipment and technologies that facilitate users to make use of m-commerce. MCD
include mobile phones, Personal Digital Assistants (PDA), portable computer notebooks, Bluetooth, WAP and other facilities that can have access to the wireless networks. Because of the need of the standardization of the application, interface and inter-connectivity of all hardware and software, it is relevant to the adoption and usage of MCD (Dholakia and Rash 2004; Buellingen and Woerter 2004). We expect that the heading towards a world of mobile networks and wireless devices, which will present a new perspective of time and space, is definitely on its way.

1.1 Objectives

Several basic questions about MCD will be addressed in this study. First, why should consumers adopt MCD? What are the influencing factors for consideration? Second, how do the MCD compare with the devices for other types of commerce such as e-commerce? Consumers will only adopt MCD when there are some potential significant advantages when comparing to old devices for other types of commerce. There is still a lack of comprehensive frameworks within which the adoption of MCD can be evaluated. Traditional viewpoints regarding this issue, especially those that are based on technology acceptance models, will need to be revisited and revised when consumers are considering such an adoption.

In this proposed research, we intend to integrate the major ideas of the technology acceptance models (both TAM and TAM2) and the options model as our basic framework for studying m-commerce. Very little research has ever been done along this direction. We can contribute to the literature by exploring and identifying the various options, or independent variables, that will affect the decision of buyers to adopt new technologies related to m-commerce. The research framework will be of interest to marketers in m-commerce and also to academics in the fields of marketing and IT. Both are keen to determine how they can perform further relevant research and position themselves well in the next generation of m-commerce. Our proposed framework will have both theoretical and practical implications through knowing why and how relevant MCD are adopted for m-commerce.

2 M-COMMERCE DEVICE ACCEPTANCE

The emergence of m-commerce requires relevant new technologies and attracts some current studies on its adoption (Xu and Gutierrez, 2006; Dholakia and Dholakia 2004; Bruner and Kumar, 2005; Okazaki, 2005; Harris et al., 2005). Empirical observation suggests that there is typically a substantial lag between the discovery of a new technology and its adoption (Doraszelski, 2004). The theories on the diffusion of innovation can be traced back to Everett Rogers (1962). Since then, many authors have worked further on this theory, but the core remained the same: when a new product/new technology is introduced, the target market can be divided into five segments along an axis of risk aversion: in the beginning there are the innovators, followed by the early adopters, the early majority, the late majority and the risk-allergic laggards. This proposed framework is not on the diffusion process but primarily on the intention to adopt new m-commerce technologies by early adopters.

Literature on the delayed acceptance of technology has stressed primarily on the benefits and use of new technologies (Davis et al., 1989), comparison between old and new technologies (Sheasley, 2000), role of sunk costs in existing technology (Salter, 1966) or in complementary technologies (Frankel, 1955). There are some models that associate diffusion lags of new technologies with the reduction of complementary costs such as specific human capital (Chari and Hopenhayn, 1990), learning-by-doing (Parente, 1994; Jovanovic & Lach, 1989), and search costs (Jovanovic & MacDonald, 1994). Other behavior-based models, such as the information cascades theory on the acceptance of new technology, suggests that an individual who adopts new technologies may do so based on the actions of others and contrary to his or her private preferences (Bikhchandi et al. 1992).

2.1 Why TAM and Options Model are used Together

In line with mainstream literature, we acknowledge that TAM will be our primary research framework. The technology acceptance model (TAM) is an information systems theory that models how users come to accept and use a new technology, with reference to two major considerations, perceived usefulness and perceived ease of use (Bagozzi et al.,
The former is about the degree to which a person believes that using a particular system will make his or her life easier, e.g. by enhancing his or her job performance or reducing the workload, while the latter is the degree to which a person believes that it is not difficult to actually use a particular system (Davis et al., 1989). An extended version of the TAM model, referred to as TAM2, was later developed to explain perceived usefulness and usage intentions in terms of social influence and cognitive instrumental processes (Venkatesh and Davis, 2000).

The options model demonstrates that a new technology with a moderate expected improvement in performance can experience substantial delays in acceptance and price dropping in a competitive market (Bessen, 1999; Sheasley, 2000). Rather than adopting a new technology that demonstrates only marginal improvement, consumers have the option of not adopting until the new technology, in terms of performance and price, is substantially better than the old technology. Consumers contemplating the adoption of a new technology are, of course, aware of the possibility of sequential improvement. They consider not only the current technical level of the new technology, but also their expectations of possible upgrades and changes in the future of the new technology (Sheasley, 2000).

While using technology acceptance models (TAM) as our primary reference, we also incorporate the important implications of the options model into our basic framework for analyzing consumers’ adoption of MCD for m-commerce. We observe that TAM is primarily about to what extent people will adopt new technologies with reference to the advantage and benefits (perceived usefulness and perceived ease of use). However, we observe that, although it is very likely that new technologies will eventually replace old ones, the devices of old and new technologies are very often being used at the same time, based on people’s assessment of the comparative merits of the two generations of technologies. We, therefore, consider that the options model (which focuses on the comparison between old and new technologies) is a useful tool in our analysis of the acceptance of MCD for m-commerce.
3 THE INTEGRATED FRAMEWORK

In essence, options model focuses on the comparison between existing and old MCD while TAM places emphasis on the generic attributes and utility of MCD. We integrate the major ideas of these two models into our new proposed model as shown in Figure 1. Based on our theoretical framework, we identify four influencing factors: merits, maturity, maneuverability, and mentality, which we consider to be relevant to the decision of consumers in adopting MCD. We also identify two generic antecedents of these influencing factors, mobility and matching. This suggests an extent of influence of these influencing factors and their antecedents to affect consumers’ adoption decision of MCD.

3.1 Influencing Factors based on Technology Acceptance Model

With reference to technology acceptance model (TAM, TAM2 and UTAUT) (Venkatesh & Davis, 2000; Venkatesh et al. 2003), we consider whether the adoption of MCD will bring advantages to consumers. We identify two M’s, maneuverability and mentality, for relating the acceptability of MCD to users.

The first influencing factor, maneuverability, is related to the perceived usefulness in the adoption of MCD and the degree to which a person can make the best use of such MCD. Consumers will tend to adopt devices that are user-friendly and do not require some intensive training of adoption (Prasanna et al, 1994). It would be measured by the usability of the MCD.

The second influencing factor, mentality, is concerned with the match between the new technology and consumers’ own mindset, as well as the appropriate recognition of their peer groups (Bessen, 1999; Venkatesh & Davis, 2000). General acceptance by the consumers, especially by their peer groups, will be very important to consumers when they consider using MCD for matching the devices of other people. Mentality can be evaluated by the perceived peer groups’ acceptance of MCD.

3.2 Influencing Factors based on Options Model

While mainstream literature on the adoption of new technologies is primarily based on the technology acceptance model, we consider that, in the context of m-commerce, we also need to think about some other aspects. With regards to the options model (Bessen, 1999; Sheasley, 2000), we consider the comparison between MCD and devices for other types of commerce, and in particular, the comparative advantages of MCD to consumers. Based on the options model, we identify two M’s, merits and maturity, in relation to the comparison.

We identify the third influencing factor, merits, which is about the degree to which a buyer believes that the MCD can provide significant improvement in the purchase process. Handheld mobile devices, such as PDA and other enhanced alphanumeric communicators, have supplemented mobile telephones, thus expanding the range of MCD available for m-commerce transactions. With the abilities to be connected to digital communication networks, MCD are considered to be in possession of important comparative advantage of mobility. Merits can be measured by the comparative advantages of the MCD in relation to the old devices for other types of commerce.

The fourth influencing factor, maturity, is the possibility that the technology of the MCD is mature enough so that there will not be any possible significant improvements at a later stage. While academic researchers and business practitioners recognize that the electronic market will penetrate and replace traditional type of commerce, there are still some reservations that will likely cause the early adopters of new technologies some problems in terms of the obsolescence of devices (Samuelsson & Dholakia, 2003). Most consumers will prefer adopting MCD with more mature technologies so that there is no need for a high level of subsequent upgrading of devices. Maturity can be assessed by the perception that the relevant MCD can or cannot be upgraded.

3.3 Generic Attributes of MCD

In addition to the identification of the influencing factors of the adoption of MCD, we also consider their antecedents, which are related to the very basic and essential characteristics of MCD. We start our analysis by considering two generic attributes of MCD, mobility and matching.

Mobility is the most fundamental aspect of m-commerce because the name m-commerce arises from the mobile nature of the wireless environment that supports mobile electronic transactions (Coursaris et al, 2003). Mobile wireless devices, such as mobile phones, PDA, and portable computer notebooks, can have the ability to help users gain
access to the Internet. Based on these wireless devices, m-commerce is a natural extension of e-commerce but can provide some additional advantages of mobility for consumers. Mobility is a major prerequisite for the adoption of MCD. It is an antecedent of the influencing factors of the adoption of MCD because people will consider adopting wireless connection because it can allow significant improvement (i.e., merits), easy to use (i.e., maneuverability), and can be accepted by peer groups (i.e., mentality). It can be measured by the extent of access to wireless networks.

Matching describes the need for the standardized and common interface of MCD (Coursaris et al., 2003). The unique characteristic of m-commerce very often requires both ends of this new type of commerce to have a common interface. M-commerce applications have the challenging task of discovering services in a dynamically changing environment. Effective mechanisms need to be in place for the interface between various types of MCD. Matching is an important antecedent of the influencing factors of consumers’ adoption of MCD because the need for standardization (i.e., matching) is important for m-commerce technology which allows for the interface of MCD with the wireless networks when the technology and interface is mature (i.e., maturity). It also provides utility for consumers for interacting with other devices (i.e., merits). Matching can be measured by the degree that MCD can be compatible with each other.

Based on our conceptual framework, we identify the various influencing factors (i.e., 4 M’s) which can affect consumers’ decision of the adoption of MCD in their purchases. It is possible to collect data on whether consumers will consider the adoption of MCD, and at the same time, researchers can also investigate the reasons why they adopt or do not adopt MCD, in terms of timing, opportunities, changing trends and applications.

4 CONCLUSIONS

We are proposing new insights and new adoption behavior in the ubiquitous world of m-commerce, which we believe, are still not yet fully understood by most marketers and scholars (Stevens & McElhill, 2000; Struss et al., 2003). Our conceptual framework contributes to literature by suggesting the new constructs: merits, maturity, maneuverability, and mentality, which we consider to be relevant to the decision of consumers in adopting MCD. It also represents an examination of the adoption of MCD by consumers in their purchase processes and will be of interest to the MCD market.

ACKNOWLEDGEMENTS

This research was supported in part by The Hong Kong Polytechnic University under grant number APa6E.

REFERENCES


IT APPLICATIONS IN PRODUCTION PLANNING AND CONTROL

A Survey of Medium Sized Business in German-speaking Europe

Jakob Lewandowski, Matthias Buhl and Burkhard Kittl
Institute for Production Engineering, Vienna University of Technology, Karlsplatz 13, 1040 Vienna, Austria
lewandowski@mail.ift.tuwien.ac.at, buhl@imw.tuwien.ac.at, kittl@mail.ift.tuwien.ac.at

Keywords: Enterprise Resource Planning (ERP), IT survey, Manufacturing Execution System (MES), Medium-sized business, Production Planning and Control.

Abstract: This paper addresses the use of IT-applications for production planning and control of medium-sized businesses in German-speaking Europe. The article summarises challenges today and expectations for the future. In particular, the kind of IT systems in use, experiences and satisfaction with IT support were key questions of the study. Basis of the research was a telephone survey with 150 IT managers of representative medium-sized enterprises in production, trade and service sectors.

1 INTRODUCTION

Production and therewith combined planning and controlling of resources are the core of every company in the manufacturing industry. Although most of the factories enforce development and adaptation of the technological processes, the capabilities of IT applications for organizing and handling the production process are still underestimated. Many troubles are arising out of linguistic barriers between suppliers/consultants and users. There is no obligatory nomenclature and the latest endeavours for standardisation are rarely used or even known by many users. This fact leads to misconceived coherences and in the worst case to prejudices regarding the qualities of IT systems.

Moreover, medium-sized enterprises don’t have sufficient IT resources compared to bigger companies. For this segment the cost-benefit ratio of automating workflows in the production planning and control level often is not in evidence. Furthermore the mid-sized business is confronted with an unclear diversity of software applications for the production area.

The topic of IT applications in production planning and control has kept many developers and scientists busy for decades but it is more than ever relevant, because there is still no satisfying solution for bridging the divide between the shop floor and management. Even the great efforts of leading business-software companies in the last years have only brought minor improvement. Getting a better insight into the use of and the opinions about IT solutions in the industry was the impulse to start a survey to collect useful information for further research projects.

2 APPROACH

The study consists of face-to-face and telephone interviews.

The subject of production planning and control is quite complex, so intensive preparation for the face-to-face interviews was indispensable. Almost half of the project time was spent on literature studies as well as developing and evaluating the questionnaire.

In the qualitative face-to-face interviews with twenty-two representative enterprises, the authors wanted to explore how comfortable the interview partners feel concerning terms and definitions in this topic. In order to get a proper data base additional telephone interviews with 150 IT managers were carried out. The point of interest was defined as systems in “Business Planning and Logistics” (layer 4) and the “Manufacturing Operations and Control” (layer 3) according to the ANSI/ISA-95 standard (Instrument Society of America, 2000).
2.1 Objectives

The basic objectives of the survey carried out in German-speaking regions were:

- Finding out the problems enterprises have in the area of production planning and control.
- Identifying and locating the use of IT solutions.
- Pin-pointing the prejudices and expectations concerning IT systems.
- Assessing the experiences and best practices with enterprise resource planning (ERP) and manufacturing execution systems (MES).

2.2 Target Group

While the European German-speaking market is dominated by small to medium sized enterprises, the businesses size chosen for the survey ranged from 50 to 500 employees. The majority of the companies belonged to the continuous or discrete producing industry, but the trade and service sector was allowed also.

2.3 Survey Model

As a basis for the survey a theoretical framework was used. It combined the tasks and functions of the “Aachener model for production and control” (Schuh, 2006) with the structure of the ANSI-ISA95 standard (Instrument Society of America, 2000). In addition, a cross-check with further American and European literature like MESA (MESA, 1997) or the VDI-report (VDI, 2006) combined the particular advantages of each model. The functional hierarchy of the ANSI/ISA-95 standard is exposed in Figure 1.

The survey model consisted of three main parts. The first one gave a hierarchical structured overview of the functions in the production and control of the manufacturing industry. In the second part the functions of order processing were drawn in a task related flow-chart. The third part is a checklist referencing functions and processes for the face-to-face interviews in order to allow free conversation and obtain structured and comparable information at the same time.

3 RESULTS

The analysis of the collected data occurred on the basis of several key questions. First of all, what kind IT systems do medium-sized enterprises use at the moment? Furthermore how satisfied are the companies with their applications? Positive and negative experiences and expectations essentially depend on the idea the enterprises have of concepts like enterprise resource planning or manufacturing execution systems. For this reason it was important to find out how conclusive these concepts are for companies. Another critical issue was to ascertain the challenges entrepreneurs are facing in the production area. Above all, how far should processes be automated respectively is it really desirable to have every process supported by an IT system in order to stay on a competitive basis?

![Figure 1: Functional hierarchy (Instrument Society of America, 2000).](image-url)
the issue of individual software becomes more coherent.

Table 1: ERP-systems in use (n=150, telephone interviews).

<table>
<thead>
<tr>
<th>Employees</th>
<th>Standard software global vendors</th>
<th>Standard software regional vendors</th>
<th>Individual software</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>In use</td>
<td>Rollout</td>
<td>Planned</td>
</tr>
<tr>
<td>Total</td>
<td>68.0%</td>
<td>2.0%</td>
<td>0.7%</td>
</tr>
<tr>
<td>9-100</td>
<td>60.5%</td>
<td>2.6%</td>
<td>1.4%</td>
</tr>
<tr>
<td>101-250</td>
<td>61.8%</td>
<td>2.9%</td>
<td>0.0%</td>
</tr>
<tr>
<td>251-500</td>
<td>87.5%</td>
<td>0.0%</td>
<td>0.0%</td>
</tr>
</tbody>
</table>

A moderate enhancement was determined relating to the rollout time of ERP-systems. Compared with studies in the last decade, the time period decreased. Concerning the fields of application like material management or production planning, there was no significant difference between standard and individual software. Material management, procurement, sales as well as controlling and accounting were implemented by most of the companies. Rare domains were production planning and quality management. The reason for this is the diversity of the manufacturing processes, even within an industry sector.

A conspicuous connection was detected between the type of ERP-system and the management of business processes. There were considerably less IT supported business processes running parallel to the ERP-system in companies using software by global vendors. It seems obvious that ERP-systems of leading vendors are employed more efficiently.

Table 2 contains a summary of the criterions leading to the decision for a certain ERP-system. The gist is that companies belonging to a concern or group usually have no bearing on decisions because most of these are predetermined. The second and third most relevant motives were flexibility and sectoral advantages. The fact that both consultancy and license fees are rated below three percent indicates that these costs are ranked as a non-determining factor regarding the total cost of ownership.

The information relating to how comfortable the medium-sized enterprises are with their IT systems can be regarded as inconspicuous, but even so significant. The data content of Table 3 gives an overview of the three groups of ERP-systems.

There was no difference in the degree of functional coverage as well as satisfaction between...
companies with standard or individual software in use. The user driven margin that individual systems naturally should have is not recognizable.

Table 2: Criterions for selecting an ERP-System (n=150, telephone interviews).

<table>
<thead>
<tr>
<th>Employees</th>
<th>Concern-directive</th>
<th>No alternatives to specific ERP</th>
<th>Known reasons</th>
<th>Consultancy fees</th>
<th>Size of vendor</th>
<th>Global application of specific ERP</th>
<th>System security</th>
<th>Vendor support</th>
<th>License fees</th>
<th>Sectoral advantage</th>
<th>Flexibility</th>
<th>N/a</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt;100</td>
<td>23.7%</td>
<td>6.6%</td>
<td>4.0%</td>
<td>2.6%</td>
<td>0.00%</td>
<td>10.5%</td>
<td>5.3%</td>
<td>1.3%</td>
<td>2.6%</td>
<td>19.7%</td>
<td>13.1%</td>
<td>18.4%</td>
</tr>
<tr>
<td>101-250</td>
<td>32.4%</td>
<td>5.98%</td>
<td>5.9%</td>
<td>0.0%</td>
<td>5.9%</td>
<td>9.9%</td>
<td>5.9%</td>
<td>2.9%</td>
<td>0.0%</td>
<td>32.3%</td>
<td>8.8%</td>
<td>11.8%</td>
</tr>
<tr>
<td>251-500</td>
<td>37.5%</td>
<td>7.5%</td>
<td>12.5%</td>
<td>2.5%</td>
<td>2.5%</td>
<td>7.5%</td>
<td>17.5%</td>
<td>0.00%</td>
<td>2.5%</td>
<td>12.5%</td>
<td>20.0%</td>
<td>12.0%</td>
</tr>
<tr>
<td>Total</td>
<td>29.3%</td>
<td>5.3%</td>
<td>4.0%</td>
<td>2.0%</td>
<td>4.7%</td>
<td>8.7%</td>
<td>8.7%</td>
<td>1.3%</td>
<td>2.0%</td>
<td>20.7%</td>
<td>14.0%</td>
<td>12.0%</td>
</tr>
</tbody>
</table>

Table 3: Functional coverage of and satisfaction with ERP-systems (n=150, telephone interviews).

<table>
<thead>
<tr>
<th>Degree of functional coverage</th>
<th>Standard software / global vendors</th>
<th>Standard software / sectoral vendors</th>
<th>Individual software</th>
</tr>
</thead>
<tbody>
<tr>
<td>Procurement / Material Management</td>
<td>91.3%</td>
<td>90.9%</td>
<td>91.0%</td>
</tr>
<tr>
<td>Sales/ Distribution/ Service</td>
<td>91.3%</td>
<td>90.4%</td>
<td>88.5%</td>
</tr>
<tr>
<td>Production-/ Project planning</td>
<td>90.3%</td>
<td>92.0%</td>
<td>92.0%</td>
</tr>
<tr>
<td>Controlling</td>
<td>93.5%</td>
<td>90.9%</td>
<td>91.8%</td>
</tr>
<tr>
<td>Quality Management</td>
<td>91.6%</td>
<td>91.6%</td>
<td>87.0%</td>
</tr>
<tr>
<td>Human Resources</td>
<td>92.7%</td>
<td>85.1%</td>
<td>90.4%</td>
</tr>
</tbody>
</table>

Changing attitudes to IT support and the adaptation of business processes were detected. In the medium-sized businesses more than 27% of the companies using software by global vendors are outsourcing their IT support. Worthy of mention is that the smaller the amount of employees of a company is the more likely IT support is outsourced. Asked for satisfaction with external IT support in general the average evaluation number was 1.82 (1=excellent, 5=insufficient).

The way medium-sized enterprises saw expectations and requirements in the near future often depended on the management. In general, companies agreed to the statement: “The more business’ processes are automated the bigger the benefit for the company”. Furthermore, companies are poised to align business processes to save costs for customizing the IT system. These two findings were emanating from the telephone interviews as well as from the face-to-face-interviews. The awareness of corporate responsibility for a successful IT implementation is obvious, the capability and benefits of ERP-systems are established. Basically there was a high degree of trust in ERP-systems.

Table 3: Functional coverage of and satisfaction with ERP-systems (n=150, telephone interviews).

<table>
<thead>
<tr>
<th>Satisfaction (1=very satisfying, 5=unsatisfied)</th>
<th>Standard software / global vendors</th>
<th>Standard software / sectoral vendors</th>
<th>Individual software</th>
</tr>
</thead>
<tbody>
<tr>
<td>Procurement / Material Management</td>
<td>1.8</td>
<td>1.5</td>
<td>1.8</td>
</tr>
<tr>
<td>Sales/ Distribution/ Service</td>
<td>1.8</td>
<td>1.6</td>
<td>1.9</td>
</tr>
<tr>
<td>Production-/ Project planning</td>
<td>1.7</td>
<td>1.8</td>
<td>1.9</td>
</tr>
<tr>
<td>Controlling</td>
<td>1.7</td>
<td>1.7</td>
<td>1.7</td>
</tr>
<tr>
<td>Quality Management</td>
<td>1.8</td>
<td>1.7</td>
<td>1.9</td>
</tr>
<tr>
<td>Human Resources</td>
<td>1.8</td>
<td>1.7</td>
<td>1.8</td>
</tr>
</tbody>
</table>

This confidence was reflected in the fact that calculating the return on investment very often is no issue to companies. They attach great importance to the factors, which are impossible to measure (e.g. transparency, communication, etc).

The expectations in the ERP-system itself rise after an adaptation phase, so long-term support is very important for the companies as well as the vendors. In most cases the area of material management indicates improvement very quickly. Users are noticing the limits of the ERP-systems in the area of manufacturing. Reported reasons are the
inflexible structure of the ERP-systems, the unsatisfying reporting capabilities and unsolved problems with machine and control interfaces. In the face-to-face interviews it was found out that IT trainings often are not mandatory in order to reduce costs during the implementing phase. At least some of the companies came to the conclusion that these retrenchments have negative effects on the system operation.

Finally the most important, because most often mentioned demands on IT software from the survey are summarised in Figure . The first four positions in the bar chart are demands for one integrated system, simple usability, central data management and flexible interfaces. Additional needs were assured vendor support, as well as monitoring and displaying the production process.

In many companies management demands more key performance indicators, this naturally militates in favour of integrated IT system. As a closing insight, companies know that there is still potential within the ERP-sector, thus it is not remarkable that the IT budgets stay equal or will increase within the next 5 years.

3.2 Manufacturing Operations and Control

Compared to the layer of business planning and logistics, the IT systems used for the shop floor show a lower level of integration respectively are not existent. Whereas almost each business (97%) uses ERP-software and the majority even works with an integrated standard software tool (more than 80%), the situation of MES-software presents itself absolutely different.

The number of enterprises, which have a totally integrated IT System is limited to approximately 10%. Most of the questioned companies apply IT systems just for certain activities in their shop floor. They use a combination of standard software tools like quality- or scheduling-software (approx. 40%), individually developed software (30%) and MS Office products (30%). About 20% of the enterprises run the production area even without any IT system.

Remarkable is the fact that IT systems in use are quite old. More than half of the companies quote to work with software which is older than ten years. Moreover there has not been any rollout of an IT system at the questioned companies in layer 3 since 2005, and the satisfaction with the existing IT applications is quite positive (with an average of 1.4 using 1 as very satisfying and 5 as unsatisfying). The companies polled in the face-to-face interviews
didn’t think it’s necessary to exchange their current IT systems in the shop floor. Some of them had already started IT projects in the shop floor, which were cancelled later due to the confusing situation of available IT solutions and vendors on the market or not evident cost-benefit ratio.

A few enterprises try to fill the gap between the layer of business planning and logistics and the manufacturing operations and control by extending their existing ERP-system. But the majority thinks that the ERP-system is not flexible enough to cope with their requirements in the shop floor, which are:

- Clustering and analysing existent data.
- Decentralising know-how.
- The need for traceability.
- The reduction of partial solutions and interfaces.
- Increased transparency.

The general requirements for IT software in the production area are shown in Figure . The main points deal with system integration and data management.

A lot of enterprises were not sure about the cost-benefit of an integrated MES solution, because there are only a few running reference systems. This could be the reason, why the integration of standard MES software is a low priority subject in many companies.

The possibility of expanding the ERP-software was more often considered by companies which are not so familiar with MES yet. Furthermore the number of employees is an indicator whether an enterprise is thinking about this alternative. The smaller a business is, the more it tries to avoid a second integrated system next to ERP. Usually the manpower of mid-sized businesses for implementation and service of a second IT system is insufficient. On the other hand there were a significant number of companies, who didn’t believe that an extension of an integrated ERP system into the shop floor is possible. They pointed out that an ERP system is not able to cope with the requirements of the shop floor. In particular flexibility and planning intervals were mentioned.

Gathering and clustering data from the shop floor is a main topic in almost every enterprise. Most of the data is transferred in the direction from layer three to four. The opposite direction regarding order data is often realized by manual instructions or documents. Only few companies, who have already dealt intensely with the subject of MES, demand a bidirectional information flow.

Although most of the enterprises don’t have an IT system for detailed scheduling, it is seen as a critical business process for the ability to respond to the market.

Another topic in the shop floor was traceability. Many companies are confronted with this subject due to legal obligations or product claims. In many cases the functionalities of ERP-solutions in this regard are not satisfying and systems without IT support are in use.

A lot of enterprises use specialised IT systems for particular functionalities in their production. The interfaces between these systems are a main, but unsolved problem. Attempts to solve this problem with Manufacturing Execution Systems were stopped, because the offered systems were not able to integrate the existing IT applications or programmable controllers satisfactorily.

Many enterprises were aware of the fact that the implementation of new systems has to go together with restructuring the processes.

Further chances of development were assigned to maintenance. This MES function turned out to be an exception, especially in the face-to-face interviews. More than 50% of end-users claimed that maintenance is a very important element but only a percentage under 10% declared to use a standard-software for planning and controlling the maintenance processes at all.

Compared to a study of Trovarit (Trovarit, 2004) the awareness level of MES has not changed significantly. More than half of the companies have not heard about MES yet. In particular enterprises with less than 350 employees usually don’t know this term. Furthermore there is a relation between the date of rollout of the ERP system and the know-how of MES. Only companies, who had applied their ERP system more than 10 years ago, have dealt intensely with the topic, or even used a MES software.

4 CONCLUSIONS

In general the potential for further development in the field of production planning and control is estimated very high. Almost each of the questioned companies quoted to struggle with one or more functions of this area. Up to the rough planning the enterprise resource planning systems got well established within the last decade. It emerged that when moving from planning to the control of the production process that the bigger part of the medium-sized enterprises places emphasis on
individual software or “no software” solutions. Although there are many old systems in use the companies are highly satisfied and prefer solutions decoupled from the ERP-system. In particular the segments quality assurance and staff work time logging are concerned.

Based on the findings of this study, one may generally conclude that for the majority of the medium-sized companies in the production industry the term MES is largely unknown. Even if some enterprises are familiar with the definition of MES they are not able to distinguish the functional environments of a manufacturing execution system. The most criticised items headed by this group are the complexity of the MES-market, as well as the unfeasible commitments from software vendors.

Also remarkable for the authors was that no enterprise polled in the face-to-face interviews had implemented a new software-system within the last three years. The statement of uncertainty relating to new software was considerable. Companies doubt that implementation of new MES will bring more benefits. The case that positive references are missing has to be mentioned additionally at this point.

Production enterprises, having long lasting experience with ERP-system are more willing to evaluate functions in the area of MES. Their know-how in implementing and maintaining this standard-software system is instrumental in realising new IT systems.

In general there is an unexplainable gap between the demands and the willingness to invest in the shop floor. In other words enterprises agreed to the advantages of an automated workflow but almost none of the polled companies has an automated workflow established or is willing to provide the required resources.

Users that exhaust the functional limits of ERP-system in the production area are increasingly claiming for more flexibility and more transparency, attributes very often expected from MES software. In the first instance the fixed framework and the not available real-time data are criticised, a standardised bi-directional interface between ERP and control layer does not exist. A high potential is attributed to this unsolved section.

Looking into the future there is an irrepealable tendency to integrated systems. The majority of the medium-sized businesses prefer one fully integrated IT system as personal resources are regularly undersized. This fact offers a chance for ERP-systems. There is still an enormous potential for the development of ERP-systems in particular concerning interfaces. These systems have to provide more and better connectivity in order to allow easy integration of satellite systems. All in all, companies are quite happy with their ERP-systems. In the sector of MES the request for one fully integrated IT system is more likely to meet the idea of an integration framework. Moreover enterprises are reluctant to substitute successful partial software solutions by a totally integrated MES. Together with the problem of low IT resources the step of exchanging the whole software landscape in the shop floor seems to be too risky for medium sized companies. But the integration of capsuled MES functions into a framework with the opportunity of step by step extension is a worthwhile alternative. Another benefit of an integration framework is that MES functionalities like maintenance can easily use existing functionality of the ERP system. Information about the production process or machine parameters are usually collected by several systems in the shop floor and the ERP. The job of the integrating framework is to allocate and analyse the information in order to provide it to a planning tool.

5 FUTURE RESEARCH

In a test facility different MES scenarios will be simulated. In one scenario it is planned to cover the functionalities of an MES by using an off-the-shelf MES product. In another scenario we are going to use an IT framework such as xMII for the integration of dedicated software tools like scheduler, shop floor data acquisition or tool management.

REFERENCES


CONDITIONS FOR TECHNOLOGY ACCEPTANCE

Broadening the Scope of Determinants of Ict Appropriation

Pieter Verdegem and Lieven De Marez
Research Group for Media & ICT (MICT), Ghent University (UGent)
Interdisciplinary Institute for Broadband Technology (IBBT), Korte Meer 7, 9000 Ghent, Belgium
Pieter.Verdegem@UGent.be, Lieven.DeMarez@UGent.be

Keywords: User research, technology acceptance, Ict appropriation, adoption determinants, usage determinants.

Abstract: Regarding the ICT industry, we have a fast evolving sector being under pressure due to a growing number of failing innovations. Companies are forced to be the first on the market and for that reason thorough insights in user preferences are indispensable. New technological innovations often fail because too much attention is given to (technical) product-related features without taking into account the most important parameters of user acceptance. In this paper we highlight some theoretical considerations on this matter. First of all, we propose an approach in which more traditional and often scattered visions on adoption determinants are broadened into an integrated framework. The approach should provide a stronger base for better targeting of (new) users of technologies. Second, we elaborate on this by rethinking these determinants with regard to the later adopters. Later adopters (or even non-users) are often ignored in technology acceptance research. However, especially for policy purposes, the understanding of why people do not adopt or do not use ICT is strongly relevant in the light of the development of the information society. Both approaches – focusing towards early as well as late adopters – are illustrated by case studies starting from a common framework.

“Consumer research has shown that knowledge about the user tends to be limited ... It is quite self-evident that both social and commercial policies will benefit from accurate insight into the different parameters determining the success (acceptance and use) of these technologies according to a user’s point of view.” (Burgelman, 2000: 236)

1 INTRODUCTION

Conditions for ‘technology acceptance’ have always been a central pillar in all kinds of approaches of studying the acceptance and appropriation of new innovations: ranging from the diffusion theory-based approaches focusing on perceived technology characteristics since the early 60’s, over more usage-oriented theoretical approaches since the 80’s to more industry-oriented studies/approaches focusing on image- and network-related determinants.

However, in today’s ICT-environment a broader and more comprehensive framework for understanding determinants or conditions for technology acceptance is more than ever needed, in order to obtain the necessary insights to face the challenges of both ICT managers and policy makers. Due to the exponentially increased offer of ICT-innovations (and as a consequence more failing technologies), all stakeholders involved are desperately seeking for accurate insights into adoption determinants as a basis for more effective introduction and targeting strategies (Lin, 1998: 95; Talukdar et al, 2002: 97; Ziamou, 2002: 366; Chen et al, 2002: 706; Venkatesh et al, 2003: 426). From a policy point-of-view such insights into drivers and barriers for adoption and usage of ICT are necessary in order to set up adequate e-inclusion measures (Chaudhuri et al, 2005: 737-739; Milner, 2006: 177; Trkman et al, 2008: 102).

In this paper we introduce a framework that could help to refine our thinking on this. First, we broaden the scope on adoption determinants by integrating the existing but fragmented approaches into a more comprehensive one. This becomes more important for industrial and marketing purposes, as a thorough understanding of the user – the customer – is necessary for acceptance. Second, we elaborate on this by paying attention to approaches that go beyond adoption diffusion. More specifically, policy makers are seeking to understand parameters that have an influence on the impact of ICT adoption and use, in order to formulate effective measures in the light of overcoming digital inequalities.
2 DETERMINANTS FOR ICT ACCEPTANCE

2.1 Broadening the Scope on Adoption Determinants

With ‘adoption determinants’ we refer to parameters that influence technology acceptance in terms of the actual adoption decision (De Marez, 2006: 189-192). For a long time and to a large extent influenced by the dominant technological deterministic paradigm, demographic variables were supposed to have an important influence on that adoption decision (see Rogers, 1983; Rogers, 2003). However, many scholars have stated that – in addition to the more traditional parameters – this view should be extended to an approach based on ‘attitudinal’ adoption determinants (Bergman et al, 1995; Plouffe et al, 2001; Atkin et al, 2003; Leung, 1998). Attitudinal determinants are related with more subjective perceptions of innovation characteristics and personality traits.

The approach of this attitudinal adoption determinants was mainly inspired by the diffusion theory, in which innovations were supposed to have a set of five characteristics (relative advantage, complexity, compatibility, trialability and observability) of which the subjective perception determines one’s attitude towards the technology, and one’s innovativeness or timing of adoption decision (Rogers, 1983; Rogers, 2003). The perception of each of these characteristics is assumed to have a strong relationship with the innovativeness of an individual. Innovators and early adopters, for example, are assumed to have a higher perception of relative advantage than the (later) majority segments, together with a lower perception of complexity of the innovation (contrary to the later adopters).

Over the years, the increasing attention paid to these ‘attitudinal’ adoption determinants resulted in a considerable yet cluttered extension of the original set of five adoption determinants. The convergence with social psychology theories such as the Theory of Reasoned Action (TRA) (Fishbein, 1967; Fishbein & Ajzen, 1975), (Decomposed) Theory of Planned Behaviour ((D)TPB) (Ajzen, 1991; Taylor & Todd, 1995) and Technology Adoption Model (TAM) (Davis, 1986; Davis, 1989) in particular led to an extremely valuable - yet fragmented - increase in (research on) adoption and determinant models. Some scholars consider one or two extra determinants (Holak & Lehmann, 1990), while others considered eight (Plouffe et al.; 2001), ten (Choi et al., 2003) or more determinants.

Downside of this increased attention is that researchers nowadays are confronted with a ‘lack of overview’, since the increased multidisciplinary interest entails a cluttered and inconveniently arranged entirety of determinants. Evidently, more accurate insight into adoption determinants requires an insight in more than the five determinants of Rogers’ diffusion theory, but it remains unclear how many and which determinants should be taken into account. Since a convenient overview of (potentially) relevant adoption determinants for ICT innovations is still lacking to date (Busselle et al, 1999; Randolph, 1999; Hadjimanolis, 2003 – an exception is the work of Venkatesh et al, 2003) we conducted a meta-analysis on determinants for ICT adoption (De Marez, 2006). Comparable to the development of UTAUT (Unified Theory of Acceptance and Use of Technology, Venkatesh et al, 2003: 446-465), we started from different studies and existing theoretical models (in the field of communication, marketing as well as social psychology) whose central building block was mainly diffusion theory’s set of five determinants. This resulted in an extension to 19 determinants, in which we distinguish ten innovation-related characteristics (perceptions), eight adopter-related characteristics, and the impact of the marketing strategy (see table 1 below).

Table 1: Extension of adoption determinants (De Marez et al, 2007: 82).

<table>
<thead>
<tr>
<th>ADOPTION DETERMINANT</th>
<th>ASSUMED RELATION WITH INNOVATIVENESS</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>INNOVATION RELATED CHARACTERISTICS</strong></td>
<td></td>
</tr>
<tr>
<td>Compatibility</td>
<td>+</td>
</tr>
<tr>
<td>Complexity</td>
<td>-</td>
</tr>
<tr>
<td>Cost</td>
<td>-</td>
</tr>
<tr>
<td>Enjoyment</td>
<td>+</td>
</tr>
<tr>
<td>Observability</td>
<td>+</td>
</tr>
<tr>
<td>Relative advantage</td>
<td>+</td>
</tr>
<tr>
<td>Reliability</td>
<td>+</td>
</tr>
<tr>
<td>Tangibles</td>
<td>+</td>
</tr>
<tr>
<td>Trialability</td>
<td>+</td>
</tr>
<tr>
<td>Visibility</td>
<td>+</td>
</tr>
<tr>
<td><strong>ADOPTER RELATED CHARACTERISTICS</strong></td>
<td></td>
</tr>
<tr>
<td>Control/ Voluntariness</td>
<td>+</td>
</tr>
<tr>
<td>Image/Prestige</td>
<td>+</td>
</tr>
<tr>
<td>Innovativeness</td>
<td>+</td>
</tr>
<tr>
<td>(product) Knowledge</td>
<td>+</td>
</tr>
<tr>
<td>Opinion leadership</td>
<td>+</td>
</tr>
<tr>
<td>Optimism</td>
<td>+</td>
</tr>
<tr>
<td>Social influence</td>
<td>+</td>
</tr>
<tr>
<td>Willingness to pay</td>
<td>+</td>
</tr>
<tr>
<td><strong>IMPACT OF MARKETING STRATEGY</strong></td>
<td></td>
</tr>
<tr>
<td>Marketing (impact)</td>
<td>+</td>
</tr>
</tbody>
</table>

Clearly, innovativeness and adoption decisions seem to be determined by more characteristics than
the original five initiated by Rogers’ diffusion theory. The perception of ‘relative advantage’ for example, can express itself in several dimensions. The ‘perceived cost’ and ‘tangibles/aesthetics’ are the most important of them. Most scholars relegate to Rogers’ work in his conceptualization of ‘observability’ in terms of the perceived result demonstrability, while some others distinguish the latter from ‘visibility’ as the degree to which the innovation is visible to others in its own right. It is also important to account for the ‘perceived enjoyment’ of using the innovation (the so-called likeability), and ‘reliability’ as a dimension of perceived risk that is not covered by other determinants (‘reliability’ in this context refers to ‘performance risk’). ‘Innovativeness’, on the other hand, is the most important personality characteristic. It covers a multitude of sub dimensions such as ‘venturesomeness’, ‘novelty seeking’, ‘cosmopolitanism’, ‘variety seeking’, ‘information seeking’, etc. ‘Opinion-leadership’ needs to be considered as a separate dimension, just as a person’s ‘optimism’ towards technology, ‘product knowledge’, ‘willingness (and ability) to pay’, the ‘perceived impact on one’s personal image’, the ‘perceived control’, ‘impact of social influences’ and the ‘impact of marketing, advertising and promotional strategies’.

If industry strategies nowadays require more profound insight in more than the traditional five determinants, it will largely boil down to an insight in these 19 determinants. It will probably never be the case that all these determinants are relevant, but if prior-to-launch research could reveal which determinants are the most important drivers and barriers for which segments, this would allow to adjust the approach of different segments. Question remains, however, how to acquire such prior-to-launch insight?

2.2 Elaboration of Determinants with Regard to ‘Later Adopters’

Another challenge of research concerning the acceptance of new technologies – especially for policy strategies – is how to gain insight in the profiles of later adopters. That are individuals to whom traditionally less attention is given in innovation studies (Selwyn, 2003: 100-101; Roe & Broos: 91). People who step later into the innovation circle or who even resist to do this, are often left aside. However, research of non or later adoption could offer fundamentally added value. First, industry or managers could learn substantially not only of why people adopt a new technology but also why they are not willing to adopt. This could provide insights in how to adjust the innovation (in all its dimensions: product, distribution, communication) in order to stimulate appropriation by the overall population. On the other hand, in view of the pervasiveness of ICT in society and the increasing dependence on ICT in everyday life, policy makers are obliged to think about policies that prevent exclusion of groups of citizens in the development of the information society. Insights in the parameters of adoption by later adopters is therefore of crucial importance.

The adoption of a certain technology (as for which the determinants are discussed in 2.1.) however, cannot be the sole focus when studying the factors that influence technology acceptance. This would be too much a technology deterministic and diffusion-based approach, mainly serving ‘industry purposes’ (how to approach the most interesting segments of innovators, early adopters, early majority as good and as soon as possible?). A more elaborated focus on technology acceptance not only requires a focus on adoption, but also on usage determinants. In addition, a thorough understanding of technology acceptance not only asks for a focus on the first segments in the diffusion curve, but also on the later segments in that curve (late majority and laggards).

Attention for digital inequalities is, both in scholarly publications as well as in political studies and in the popular press and media is, an obvious result of the euphoric ‘cyberbole’ that characterized much of the rhetoric of new technologies since the mid-1980’s (Gunkel, 2003: 500). Hence, profound insights in why people lag behind in the adoption and use of new technologies, are important in view of the development of the information society for all. More insights are necessary, especially when we can conclude that business strategies and policies that were successful in, for instance, increasing internet penetration in the early days, may no longer be appropriate to reach the rest of the society. And this is most probably so in societies where a majority of people are already connected to the internet. Thus, policies also need insights in the most important drivers and barriers that have an impact on the individual’s decision to appropriate an ICT product.
3 CASE STUDIES

3.1 Broadening for Industry & Marketing Purposes

The above-mentioned question was also the central question in two recent case studies conducted by Research Group MICT-IBBT. Both studies were set up to acquire the necessary insights in attitudinal adoption determinants for two ‘mobile innovations’, in preparation for their commercial launch in Belgium. In the first case-study (2006) a sample of 269 respondents was questioned about their attitudes towards a new ‘mobile news’-application in the context of the IBBT-project ROMAS. In the second case study (2007) a representative Flemish sample of 405 respondents was questioned about their attitudes towards mobile television services. In the first study, data were collected by means of an online survey (after a two months period in which the respondents could test the mobile news application). In the second study, data were collected by means of 40 minutes during CAPI-interviews (in which respondents were shown short movies on DVD in order to familiarize them with mobile tv applications and usage moments). In both studies potential adopter segments (innovators up to laggards) for the innovations were forecasted by means of the Product Specific Adoption Potential scale (De Marez, Verleye, 2004a,b), and the 19 determinants were transformed into a battery of 47 Likert statements (cf. table 2), to be answered on 5-point agreement scales (varying from 1: ‘I do not agree at all’ to 5: ‘I fully agree’).

Table 2: Operationalisation of determinants in 47 Likert statements (applied to the mobile news/TV cases).

<table>
<thead>
<tr>
<th>Determinant</th>
<th>Likert Scale</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Subscription on mobile news/TV services seems expensive to me.</td>
<td></td>
</tr>
<tr>
<td>2. Mobile news/TV services seem expensive to me.</td>
<td></td>
</tr>
<tr>
<td>3. Mobile news/TV services will probably cost too much for many people.</td>
<td></td>
</tr>
<tr>
<td>4. Mobile news/TV service seems very user friendly to me.</td>
<td></td>
</tr>
<tr>
<td>5. Mobile news/TV services will certainly make some things easier for me.</td>
<td></td>
</tr>
<tr>
<td>6. I think to be among the first to subscribe to such mobile news/TV services.</td>
<td></td>
</tr>
<tr>
<td>7. I do not see where or when to use mobile news/TV services.</td>
<td></td>
</tr>
<tr>
<td>8. I fear that mobile news/TV services application offers different possibilities, which makes it rather complicated.</td>
<td></td>
</tr>
<tr>
<td>9. If I buy a new mobile, it has to be a model that fits my personality</td>
<td></td>
</tr>
<tr>
<td>10. If I would mind if that would imply an investment in a new device.</td>
<td></td>
</tr>
<tr>
<td>11. The advantages of mobile news/TV services are clearer to me.</td>
<td></td>
</tr>
<tr>
<td>12. One of the nice things of a mobile news/TV application is that it is something to show off with among friends.</td>
<td></td>
</tr>
<tr>
<td>13. Mobile news/TV services are only interesting to me as a part of the subscription on other mobile services.</td>
<td></td>
</tr>
<tr>
<td>14. Consultation of mobile news/TV services fits my lifestyle</td>
<td></td>
</tr>
<tr>
<td>15. If mobile news/TV would be introduced on the market, people in my environment who use mobile news/TV services.</td>
<td></td>
</tr>
<tr>
<td>16. I see many people in my environment who use mobile news/TV services.</td>
<td></td>
</tr>
<tr>
<td>17. I recently send something about mobile news/TV services</td>
<td></td>
</tr>
<tr>
<td>18. I fear that subscribing to a mobile news/TV application would be way above my budget.</td>
<td></td>
</tr>
<tr>
<td>19. I recently send something about mobile news/TV services and recently talked to someone about it.</td>
<td></td>
</tr>
<tr>
<td>20. The mobile news services application seems very user-friendly to me.</td>
<td></td>
</tr>
<tr>
<td>21. If I would use mobile news/TV services, people in my environment would look odd at me.</td>
<td></td>
</tr>
<tr>
<td>22. If I would have to use such mobile news/TV applications on my own, I don’t think I would manage.</td>
<td></td>
</tr>
<tr>
<td>23. I recently send something about mobile news/TV services and recently talked to someone about it.</td>
<td></td>
</tr>
<tr>
<td>24. I am perfectly able to explain the strengths and the weaknesses of mobile news/TV services to others</td>
<td></td>
</tr>
<tr>
<td>25. I would consider mobile news/TV adoption, I would first check the ads, brochures and promotions.</td>
<td></td>
</tr>
<tr>
<td>26. If I would subscribe to a mobile news/TV application, it would be important to me that it is provided by a well-known ‘brand’.</td>
<td></td>
</tr>
<tr>
<td>27. If I would consider mobile news/TV adoption, I would first check the ads, brochures and promotions.</td>
<td></td>
</tr>
<tr>
<td>28. If I would subscribe to a mobile news/TV application, it would be important to me that it is provided by a well-known ‘brand’.</td>
<td></td>
</tr>
<tr>
<td>29. The mobile news/TV services application offers different possibilities, which makes it rather complicated.</td>
<td></td>
</tr>
<tr>
<td>30. Mobile news/TV services are only interesting to me as a part of the subscription on other mobile services.</td>
<td></td>
</tr>
<tr>
<td>31. Mobile news/TV services are only interesting to me as a part of the subscription on other mobile services.</td>
<td></td>
</tr>
<tr>
<td>32. Subscribing to mobile news/TV services beams out a certain standing.</td>
<td></td>
</tr>
<tr>
<td>33. Subscribing to mobile news/TV services applications would have a positive impact on my image and social status.</td>
<td></td>
</tr>
<tr>
<td>34. Based on what I already knew about the application and what I have learned today, I will certainly search for more information about subscribing to these services.</td>
<td></td>
</tr>
<tr>
<td>35. I consider myself well-informed about the possibilities and (dis)advantages of mobile news/TV services.</td>
<td></td>
</tr>
<tr>
<td>36. Mobile news/TV services will certainly make some things easier for me.</td>
<td></td>
</tr>
<tr>
<td>37. Mobile news/TV services seem expensive to me.</td>
<td></td>
</tr>
<tr>
<td>38. Mobile news/TV service seems very user friendly to me.</td>
<td></td>
</tr>
<tr>
<td>39. Optimism</td>
<td></td>
</tr>
<tr>
<td>40. The fast technological developments are a good thing.</td>
<td></td>
</tr>
<tr>
<td>41. If you don’t want to run behind, adoption of new technologies is necessary.</td>
<td></td>
</tr>
<tr>
<td>42. I fear that subscribing to a mobile news/TV application would be way above my budget.</td>
<td></td>
</tr>
<tr>
<td>43. I fear that subscribing to a mobile news/TV application would be way above my budget.</td>
<td></td>
</tr>
<tr>
<td>44. Mobile news/TV service seems very user friendly to me.</td>
<td></td>
</tr>
<tr>
<td>45. Perceived risk (implementation)</td>
<td></td>
</tr>
<tr>
<td>46. If I have no problem to sort out on myself how mobile news/TV application work and must be installed.</td>
<td></td>
</tr>
<tr>
<td>47. Enjoyment</td>
<td></td>
</tr>
<tr>
<td>48. I do not see where or when to use mobile news/TV services.</td>
<td></td>
</tr>
<tr>
<td>49. I find it difficult to see other people using mobile news/TV services</td>
<td></td>
</tr>
<tr>
<td>50. I do not see where or when to use mobile news/TV services.</td>
<td></td>
</tr>
</tbody>
</table>

295
Table 2: Operationalisation of determinants in 47 Likert statements (applied to the mobile news/TV cases) (cont.).

| RELATIVE ADVANTAGE | 11. The advantages of mobile news/TV services are clearer to me than the disadvantages. |
| RELIABILITY        | 31. I doubt the reliability and proper functioning of the mobile news/TV services application. |
| SOCIAL INFLUENCE   | 9. Most people in my environment will certainly be enthusiast about the mobile news/TV application. |
| SOCIAL INFLUENCE   | 2. If ‘mobile news/TV usage’ would be considered as ‘trendy’ in my environment, I would certainly consider subscribing to it. |
| SOCIAL INFLUENCE   | 10. My direct environment will probably expect me to be one of the first to use mobile news/TV services. |
| SOCIAL INFLUENCE   | 32. Even if I am interested, I would not subscribe if my environment would be negative about mobile news/TV applications. |
| SOCIAL INFLUENCE   | 47. If I would use mobile news/TV services, it would certainly tell something about me and my personality. |
| SOCIAL INFLUENCE   | 3. Before subscribing to a mobile news/TV application, I would like the advice of some people. |
| SOCIAL INFLUENCE   | 16. Mobile news/TV services will certainly be a topic of discussion among my friends and family. |
| SOCIAL INFLUENCE   | 37. I am interested in subscribing to mobile news/TV services, but only if there are sufficient people in my direct environment doing so. Otherwise, the application wouldn’t have much value to me. |
| TANGIBILITIES      | 25. As the mobile news/TV services is presented and testable now it has an attractive design and style. |
| TANGIBILITIES      | 14. If I would consider buying a new mobile, design would be a very important buying argument to me. |
| TRIALABILITY      | 41. I would like to try out mobile news/TV services before subscribing to them. |
| TRIALABILITY      | 28. Before subscribing or adopting mobile news/TV services I prefer to look around for a while and see how others are experiencing the application. |
| VOLUNTARINESS     | 42. If I would subscribe to mobile news/TV services, it would completely be my own decision. No one would influence me in making that decision. |
| WILLINGNESS-TO-PAY| 22. Even if it costs a bit more, mobile news/TV is something I really want. |

The transformation of determinants into a scale of 47 items is the combined result of desk research and qualitative research by means of focus group interviews. A first phase of desk research resulted in a long list of statements of 19 determinants used in other studies and models (both diffusion theory based models as well as social psychology based models). In addition to this, the long list was verified in four focus group interviews with the goal to select the best way to translate the item into a statement.

All 269 (mobile news study) and 405 (mobile television study) respondents completed the entire questionnaire. The most important results show a striking difference between the attitudes or determinants for both innovations. In the average agreement scores, for example, it can be noticed that a determinant as ‘tangibles’ (14, 25) is more important for mobile television than for mobile news. Regarding ‘reliability’ (31) people seem to be more sceptical for mobile news, while the ‘perceived control’ (46) seems to result in a higher score for this new mobile application. ‘Product knowledge’ (19, 35) on the other hand is lower for mobile television; etc …. With an $R^2$ ranging between .503 and .795 for the earlier adopters and early majority, these 47 ‘determinant operationalisations’ certainly seem to be a good set of variables to explain the variance in the dependent variable ‘adoption intention’. Even for the later adopter segments this $R^2$ still ranges between .34 and .42. Detailed information about the psychometric reliability and validity can be found in De Marez et al (2007: 86-88).

Thus, for both technological innovations, this set of attitudinal determinant statements explains adoption intentions quite well, but there remain many differences in the significant determinants for the different innovations and adopter segments. ‘Lifestyle compatibility’ (39) for example is only significant for the mobile television’s innovators, not for mobile news. Also the ‘cost perception’ (1) is only significant in the mobile television case, whereas for mobile news, the ‘cost perception’ (1) is only significant in the mobile television case (laggards). ‘Triability’ (41) then is significant in both cases, but not for the same segments. Other determinants such as the perceived impact of adoption on one’s ‘image’ (33) was only significant for mobile news’ innovators and laggards. So, we can notice many differences in attitudes, as well when compared over the two cases, as compared over the different adopter segments. This emphasizes the need for a product- and segment specific approach when studying adoption determinants.

3.2 Elaboration for Policy Purposes: Analysis of Non-Adopters

The need for more profound insights in why people do not use ICT innovations, for instance computer and internet, is an important question for policy makers. For instance, as more people are connected and taking full advantage of new possibilities that are offered via internet, government cannot ignore
those groups that are not yet connected. So, policy makers should at least take the initiative to set up measures that can help people – who risk to be excluded – to enhance their participation in the information society.

Research Group MICT-IBBT was commissioned with this research question by Fedict (The federal public agency for information and communication technology) of the Belgian federal government. As the responsible agency for stimulating ICT acceptance and use in society, Fedict needed a scientific supported base for setting up new initiatives. The research results presented below draw on the experience that the Belgian government has acquired through the ‘Internet for all’ project in 2006. The latter was set up in collaboration with ISPs, PC manufactures and retailers, and entailed the provision of an affordable package (€750 - €1000) deal to customers, consisting of a PC, an internet connection plus a training session. It was calculated that the project contributed to 16% of the increase of new internet connections over a period of one year (Verdegem & Verhoest, 2008: 38). A critical evaluation of the ‘Internet for all’ project revealed different elements, two of which inspired our research. The first was merely the confirmation of what could be expected. Not all of the groups in society were equally well served by the campaign. For instance, for some individuals the proposed offering was too expensive. The second source of inspiration was an incidental call of a representative of a professional organization of physical therapists that proposed to target the campaign also towards the members of his organization. These two observations triggered a reflection that inspired the new policy approach and adjoining research.

The new approach is articulated around the concept of ‘relative utility’, a sociological reinterpretation of the economic concept of ‘marginal utility’. Contrary to the other case studies illustrated in this paper, of which the goal was to broaden the insights concerning adoption determinants, this case is focused on the elaboration and interpretation of parameters of ICT appropriation. By paying attention to both the adoption as well as the usage decision we wanted to provide input for measures that would help to stimulate ICT adoption and use. Following the relative utility approach, the assumption is that the specific combination of conditions in terms of access to ICT, skills to master the devices and attitudes towards the technology, has an impact on whether people will use ICT or not. More specifically, based on the combination of perceptions of people towards access, skills and attitudes (ASA) it becomes possible to determine a hypothetical ‘turning point’ for ICT use, namely the point at which the benefits will outweigh the cost of appropriating an ICT product for a certain category of users.

On a practical level, in order to set up effective e-inclusion measures, the advantage of this method is that groups of individuals with relatively homogeneous ASA-profiles, can easily be identified and reached by policy makers. Very often they are represented by professional or social organizations that know how to reach them and are willing to cooperate with government. A specific offering can then be proposed to these groups, taking into account the specificities of their ASA-profile and socio-economic background.

The approach draws upon the assumption that members of socio-demographically and socio-economically homogeneous groups yield similar perception in terms of access, skills and attitudes towards ICT. This hypothesis was tested by means of a quantitative survey (personal interviews with 184 respondents). The research population was composed of a theoretical sampling, meaning that we selected individuals based on a limited number of characteristics, i.e. variables of which previous research has shown that they are of major importance for (non-)adoption of ICT. In the research we recruited individuals (non-users) from ten groups, varying from single mothers with children to physical therapists. This resulted in certain prototypical profiles, exemplary for the societal diversity without being representative for the overall population (for detailed information see Verdegem & Verhoest, 2008).

In order to map the respondents’ perceptions of computer and internet use at home, we presented them with a list of statements. The statements were based on the same adoption determinants that are mentioned above (see table 2). A number of these statements aimed at obtaining information about the respondent’s specific ASA-profile: 1) positive or negative attitudes towards computer and internet at home; 2) the presence or lack of skills and competences towards using ICT and 3) the presence or absence or barriers to access ICT. Other statements served as measurement scales to gain insights in more generic factors such as, for example, the influence of social networks or marketing strategies of the ICT industry.

Based on the answers of the respondents on the statements cluster analysis revealed five distinctive groups of domestic non-users of computer and internet:

- Incapable refusers;
- Self-conscious indifferenters;
• The willing but incapable;
• Skilled ICT-lovers with limited access;
• Price sensitive pragmatists.

The clusters demonstrate that non-adopters or non-users should not be seen as one generic group. Each profile represents a different combination of the factors investigated, in which each factor carries a different weight. Statistical testing was also conclusive about the relationship between the group membership (from theoretical sampling) and the membership of the ASA-profiles (Pearson Chi-Square $p \leq 0.01$). As such, we found empirical foundation for the assumption that homogeneous groups – in terms of socio-demographic and socio-economic characteristics – result in generic ASA-profiles.

Following on this quantitative research qualitative in-depth interviews and focus group interviews were organized to refine our thinking of why people do not use ICT and to examine which levers could lift them over the turning point between non-usage and usage. The results of both research stages show the advantage of the approach proposed and offers the opportunity for policy makers to set up measures to stimulate later adopters to ICT appropriation. These measures could be more effective as they are based on strategies of segmentation and differentiation, taking into account the different profiles of these individuals. The elaboration of adoption and usage determinants is thus necessary to gain insight in a group of individuals that are often ignored in innovation research.

4 CONCLUSIONS

Our research results clearly show the need for a thorough understanding of user attitudes towards ICT acceptance. As more technological innovations are introduced in rapid succession and an increased number of those innovations is failing, accurate insights in the determinants towards adoption and use become increasingly important. We could state that both our theoretical reconsiderations as well as the empirical foundations of them, provide ICT managers as well as policy makers with useful input in support of their innovation strategies. As a matter of fact, the development of an information society for all serves both economic as social purposes.

The approach proposed started from the same common framework, i.e. more traditional adoption determinants who are founded by technological deterministic inspired paradigms. We illustrated that these parameters, who have an impact on technology acceptance, should be reconsidered. The described elaboration contains both an exercise of broadening and deepening.

First of all, it is important to examine which determinants are of major relevance in order to forecast how new innovations should be brought to the market to persuade the potential (first) adopters, or those interested in the product. Not only the product development in terms but also the targeting and marketing campaigns strongly ask for accurate insights into user preferences. Particularly in the (pre-)launch phase.

In addition, the framework of adoption determinants should also be re-evaluated with regard to later adopters. People who enter the adoption process in a later stadium – or who even resist to adopt – may have clear reasons for that. However, deep understanding of who is making less (or even no) use of information technologies remains weak. Nevertheless, this is of major importance for both policy makers, as well as for ICT managers.

So, in a nutshell, our approach contains both a managerial as a policy relevance. Furthermore, we also hope that this paper contributes to both theoretical reconsiderations as well as the methodological foundation of technology acceptance research.

ACKNOWLEDGEMENTS

This work was supported by the IBBT projects ROMAS (Research on Mobile Applications & Services) and MADUF (Maximizing DVB Usage in Flanders). Both projects are funded by the Interdisciplinary Institute for Broadband Technology (IBBT) and a consortium of companies. The last case study is funded by the Federal Public Agency for Information and Communication Technology (Fedict – Belgian Government).

REFERENCES

voorzieving (277-325). Fatima Reeks, Amsterdam: Meboprint.


Keywords: Innovation resistance, Adoption, Mobile phone, Banking.

Abstract: The objective of this study is to explore barriers to mobile banking adoption in two distinct European countries namely Finland and Portugal. Even successful innovation may face various types of resistance that may paralyse customers' desire to adopt or use the innovation. We investigated the country effect to five adoption barriers namely usage, value, risk, tradition and image, derived from the earlier literature. An Internet questionnaire was developed and 3,597 usable responses were collected. A confirmatory factor analysis was implemented with SEM to build the constructs' latent score levels. Using non-parametric difference tests we concluded that the resistance is significantly lower among the Portuguese online bank customers in terms of four out of the five barriers. The results can be used for a better understanding and enhancement of adoption of this specific case of m-commerce.

1 INTRODUCTION

The wide penetration and rapid diffusion of mobile phones has opened opportunities for new innovations in the services sector. One such innovation is mobile banking representing one of the most promising, while still marginally adopted, m-service. Previous studies have shown that mobile banking increases efficiency and convenience in bill paying, for example, as the service can be used wherever wanted enabling time savings and immediate reactions to unexpected service need (Laukkanen and Lauronen, 2005; Laukkanen, 2007a).

Finland has long been seen as the most successful European country in terms of the adoption and use of mobile services (Bouwman et al., 2007). However, even though already around two thirds of the Finns pay their bills over the Internet, mobile banking has not yet received the attention of the masses. In general, Finland is referred as one of the leading European countries in terms of Internet banking adoption, while, for example, Portugal is lacking far behind (Eurostat, 2007). In this study we investigated what inhibits mobile banking adoption in these two European countries and how the countries differ in terms of barriers to the service adoption.

First we describe the Internet and mobile communications market both in Finland and Portugal. Thereafter, we summarise the relevant literature on innovation resistance and banking technologies. Finally, the findings are presented and concluding remarks drawn.

2 INTERNET AND MOBILE COMMUNICATIONS MARKET: FINLAND VS. PORTUGAL

During the last decade the penetration of mobile phones has been dramatic. In addition, the diffusion of Internet-connected computers has been remarkable in the 21st century. These advances in communication technologies have reshaped the service development and revolutionised the service consumption. In Finland the amount of Internet-connected computers per 100 persons have grown from roughly 20 in 2002 to nearly 50 in 2006...
Compared to Portugal, the relative amount of these devices is over three times higher in Finland. These figures may partly explain the higher Internet banking adoption rates in Finland, even though the growth rate of these devices in Portugal has been dramatic during the last years.

![Figure 1: Internet-connected computers per 100 persons (Statistics Finland, 2008).](image1)

Although the distribution and penetration of mobile handsets in Finland is argued to be among the highest in the world, making the country an interesting test-market for new mobile services (Bouwman et al., 2007), the number of mobile phone subscriptions is even higher in Portugal with 1.16 connections per capita compared to 1.08 in Finland. The relatively low number of computers connected to the Internet and a great number of mobile phones make Portugal a highly potential market for mobile services such as banking.

![Figure 2: Mobile telephone subscription per capita (Statistics Finland, 2008).](image2)

The future of mobile communications relies heavily on services. However, the optimistic and experimental mood that we witnessed in the beginning of the century has been replaced by a cautious atmosphere in which fewer risks are taken in the development and marketing of new third generation mobile services (Bouwman et al., 2007). Therefore, insight into the reasons why consumers are not adopting mobile services is needed. In this paper we explore the adoption barriers to mobile banking in the light of consumer resistance to innovations.

3 LITERATURE

Albeit the pro-innovation bias (Sheth, 1981; Ram, 1987; Rogers, 2003) that majority of the diffusion literature has, there may be product and service categories or market segments where innovation resistance is predominant (Gatignon and Robertson, 1991). The literature on innovation resistance aims to explore the various reasons that inhibit innovation adoption. Sheth (1981) suggests that the two key factors explaining the phenomenon are habit or satisfaction with an existing behaviour and perceived risks associated with innovation adoption. He states that the inclination toward an existing behaviour is related to the typical human tendency to strive for consistency and status quo rather than to continuously search for new behaviours.

Consequently Ellen et al. (1991), note that satisfaction with current performance increases resistance to alternatives and reduces the likelihood of adoption. They further highlight the role of perceived self-efficacy which means the perceived ability or skill to successfully perform a given task. These lines of thought assume that consumers base their decisions on two aspects: perceived benefits over existing methods and perceived risks associated with innovation adoption.

Ram and Sheth (1989) suggest a more comprehensive view on innovation resistance by explaining the phenomenon with five adoption barriers namely usage, value, risk, tradition and image.

3.1 Usage Barrier

Ram and Sheth (1989) suggest that the usage barrier relates to the situation in which an innovation is not compatible with existing workflows, practices or habits. In the context of technological innovations, however, this construct parallels with complexity which, according to Rogers (2003), refers to the degree to which an individual considers an innovation to be relatively difficult to understand and use.

The small size of mobile devices including small screens and tiny multifunction keypads may be troublesome to use and hamper the usability of mobile services. Earlier studies on mobile banking show that the smaller screens appear adequate in
information-based mobile services, such as request for account balance service, but those banking services that involve transactions require a bigger screen size (Laukkanen, 2007b). For example, some bank customers consider bill payment via mobile handheld device to be difficult and time consuming as the device enables only a limited amount of information processing and for this reason, the whole bill is not visible on the display inhibiting the progress in the service process (Laukkanen and Lauronen, 2005; Laukkanen, 2007a). Moreover, some studies highlight the importance of simple authorization mechanisms in mobile banking (Laukkanen and Lauronen, 2005) while some report inconvenience due to changing PIN codes among some bank customers as the codes need to be carried along (Kuisma et al., 2007).

3.2 Value Barrier

The degree to which an individual believes that an innovation is better than the idea it supersedes determines the individual's decision to use the innovation (Rogers, 2003). This is called relative advantage which is a related concept with the value barrier referring to the performance and monetary value of an innovation in comparison to its substitutes (Ram and Sheth, 1989).

In similar vein, the greater the perceived advantage that mobile banking offers over other ways of banking, the more likely it is to be adopted (Brown et al., 2003). The earlier studies show that the option to check the movements or transactions of an account wherever wanted increases customers' feeling of control over their financial affairs adding value to service consumption (Laukkanen and Lauronen, 2005).

However, if an innovation does not offer greater performance to existing alternatives, it is not worthwhile for consumers to change their behaviour (Ram and Sheth, 1989). The extent to which an individual believes that using mobile banking is uneconomical, for instance, has a negative effect on the intention to use mobile banking (Luarn and Lin, 2005).

3.3 Risk Barrier

The risk perceptions in technological innovations usually arise due to the uncertainty to the technology's capability to deliver its expected outcome (Im et al, 2008). Thus, the diffusion of innovation is likely to take the longer the more risk adverse the innovation is (Dunphy and Herbig, 1995).

As with many other technological innovations, there appear to be security and privacy concerns to mobile banking among some bank customers (Luarn and Lin, 2005). Safety measures of personal details and financial information by the bank are one of the critical factors for the commercial success of mobile banking (Brown et al., 2003). A portable list of PIN codes may also pose security threats as it may be lost by a customer and found by an untrustworthy party (Kuisma et al., 2007).

Moreover, the extent to which a person believes a new technology will perform a job consistently and accurately (i.e. reliability) is highly important risk-related factor in technology-based financial service innovations (Lee et al., 2003). Mobile phones, for instance, may be limited in computational power, memory capacity and battery life, limiting the use of mobile services (Siau and Shen, 2003).

3.4 Tradition Barrier

The tradition barrier is related to the change an innovation may cause in a consumer's daily routines. Thus, if the consumer considers routines important in his/her daily behaviour, the tradition barrier will most likely be high. Moreover, the tradition barrier may arise when an innovation is incompatible with the consumer's existing values, norms and past experience (Ram and Sheth, 1989). Thus, an innovation needs to be well-suited with the existing values and norms in order an individual to adopt the innovation (Rogers, 2003).

Kuisma et al. (2007) showed that some consumers resist Internet banking due to their habit of paying bills via bill paying ATMs. Alternatively, a customer may need social interaction and enjoy talking to bank personnel as a strong desire to deal with human tellers is found to discourage consumer from adopting self-service technologies in banking (Marr and Prendergast, 1993). Thus, it may be that in mobile banking the tradition barrier arises if an individual simply prefers to deal directly with the bank clerk instead of using new banking technologies.

3.5 Image Barrier

The image barrier arises from unfavourable associations to the identity of the innovation, such as the country of origin, brand or the product category to which the innovation belongs (Ram and Sheth,
1989). In the case of technological innovations, for instance, image barrier may derive from a negative image of new technology in general and of a product class in particular.

In the late 90’s Fain and Roberts (1997) argued that the image barrier in online banking derives from a negative hard-to-use image of computers and the Internet. We argue that this may well be the case in mobile banking today as some consumers may perceive the mobile technology to be too difficult to use and therefore instantly form a negative image of the service related to the mobile technology.

3.6 Hypotheses Development

Following the earlier literature on innovation resistance a research model was designed (Figure 3). According to the Eurostat’s (2007) statistics, Finland is among the leading European countries in terms of individuals’ Internet banking adoption with 63 percent adoption rate in 2006. Portugal, for example, represents the opposite with only 10 percent adoption rate. Based on these facts we hypothesise that the resistance to electronic banking services, including mobile banking, is significantly lower among the Finns compared to Portuguese bank customers. This leads us to the following hypotheses:

H1: Usage barrier to mobile banking is significantly lower among the Finns compared to Portuguese

H2: Value barrier to mobile banking is significantly lower among the Finns compared to Portuguese

H3: Risk barrier to mobile banking is significantly lower among the Finns compared to Portuguese

H4: Tradition barrier to mobile banking is significantly lower among the Finns compared to Portuguese

H5: Image barrier to mobile banking is significantly lower among the Finns compared to Portuguese

H6: Overall resistance to mobile banking is significantly lower among the Finns compared to Portuguese

4 DATA AND METHODS

Based on the theory of innovation resistance and the existing literature on banking technologies, especially on mobile banking, a survey questionnaire was designed. The five adoption barriers were examined with 17 statements expressed in Table 1. A seven-point Likert scale ranging from totally disagree (1) to totally agree (7) was used.

<table>
<thead>
<tr>
<th>Statements measuring the barriers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Usage barrier</td>
</tr>
<tr>
<td>B1. In my opinion, mobile banking services are easy to use *</td>
</tr>
<tr>
<td>B2. In my opinion, the use of mobile banking services is convenient *</td>
</tr>
<tr>
<td>B3. In my opinion, mobile banking services are fast to use *</td>
</tr>
<tr>
<td>B4. In my opinion, progress in mobile banking services is clear *</td>
</tr>
<tr>
<td>B5. The use of changing PIN codes in mobile banking services is convenient *</td>
</tr>
<tr>
<td>Value barrier</td>
</tr>
<tr>
<td>B6. The use of mobile banking services is economical *</td>
</tr>
<tr>
<td>B7. In my opinion, mobile banking does not offer any advantage compared to handling my financial matters in other ways</td>
</tr>
<tr>
<td>B8. In my opinion, the use of mobile banking services increases my ability to control my financial matters by myself *</td>
</tr>
<tr>
<td>Risk barrier</td>
</tr>
<tr>
<td>B9. I fear that while I am paying a bill by mobile phone, I might make mistakes since the correctness of the inputted information is difficult to check from the screen</td>
</tr>
<tr>
<td>B10. I fear that while I am using mobile banking services, the battery of the mobile phone will run out or the connection will otherwise be lost</td>
</tr>
<tr>
<td>B11. I fear that while I am using a mobile banking service, I might tap out the information of the bill wrongly</td>
</tr>
<tr>
<td>B12. I fear that the list of PIN codes may be lost and end up in the wrong hands</td>
</tr>
<tr>
<td>B13. I trust that while I am using mobile banking services, third parties are not able to use my account or see my account information *</td>
</tr>
<tr>
<td>Tradition barrier</td>
</tr>
<tr>
<td>B14. Patronizing in the banking office and chatting with the teller is a nice occasion on a weekday</td>
</tr>
<tr>
<td>B15. I find self-service alternatives more pleasant than personal customer service *</td>
</tr>
<tr>
<td>Image barrier</td>
</tr>
<tr>
<td>B16. In my opinion, new technology is often too complicated to be useful</td>
</tr>
<tr>
<td>B17. I have such an image that mobile banking services are difficult to use</td>
</tr>
</tbody>
</table>

*Reversed scale
The questionnaire was first designed in Finnish and thereafter translated to English. The English questionnaire was then translated to Portuguese. The questionnaires both in Finland and Portugal were placed in a log-out page of large banks’ online service. Due to a vast number of online banking users in Finland the questionnaire was open much longer in Portugal than in Finland. In Finland the questionnaire was open for 72 hours between November 6th and 9th 2006, whereas in Portugal the questionnaire was open for 2 weeks, between June 28th and July 13th 2007. The surveys generated a total random sample of 3597 usable responses without missing values.

The Finnish sample is slightly male dominated (53%) and relatively young with 36.7 percent of the respondents being less than 35 years old. Finnish sample consisted of a total number of 1,494 valid responses of which 28 percent (419 cases) represented mobile banking users. The Portuguese sample is largely male dominated (61%) and even younger than the Finnish sample with 59.3 percent of the respondents being less than 35 years old. A total number of 2,103 valid responses were obtained from Portugal with 32.7 percent (688 valid cases) of mobile banking users.

In the data analysis phase the scales of positively formed statements were reversed so that the scales of all statements were comparable. Thus, a higher mean of a statement determines higher resistance of the respondent. A Structural Equation Model (with AMOS 7.0 software) was estimated and its fit and constructs’ reliability was checked. Latent scores were also computed at the constructs’ level. Using non-parametric tests (Kolmogorov-Smirnov Z and Mann-Whitney U), the differences between the countries were assessed (variables and latent scores did not follow normal distribution).

5 RESULTS

The constructs’ Cronbach’s Alphas indicated satisfactory internal consistency reliability (usage=0.92; value=0.60; risk=0.80; tradition=0.59; image=0.65) and the estimated structural model showed an acceptable fit ($\chi^2=2896.10$; d.f.=115; p=0.00; CFI=0.90; RMSEA=0.08; GFI=0.91). All coefficients revealed to be significant.

The second order structural model showed that the resistance to the adoption of mobile services mainly derives from usage (standardized structural impact=0.93) and value (0.84) barriers. Image and risk barriers also influenced the overall resistance, the effects being 0.51 and 0.31 respectively. On the other hand, tradition had a negative influence to the overall resistance, with a standardized structural coefficient of −0.15.

The latent scores were calculated using Kolmogorov-Smirnov Z and Mann-Whitney U tests. Both tests showed statistically significant differences to the five constructs and overall resistance level (Table 2). Apart from the tradition barrier, the results indicated higher values for Finland in all the barriers explored.

Table 2: Resistance levels across countries (latent scores).

<table>
<thead>
<tr>
<th></th>
<th>Finland</th>
<th>Portugal</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall resistance</td>
<td>4.052</td>
<td>3.516</td>
<td>0.000</td>
</tr>
<tr>
<td>Value barrier</td>
<td>3.058</td>
<td>2.781</td>
<td>0.000</td>
</tr>
<tr>
<td>Image barrier</td>
<td>2.345</td>
<td>2.159</td>
<td>0.000</td>
</tr>
<tr>
<td>Tradition barrier</td>
<td>3.555</td>
<td>4.047</td>
<td>0.000</td>
</tr>
<tr>
<td>Risk barrier</td>
<td>3.879</td>
<td>3.594</td>
<td>0.000</td>
</tr>
<tr>
<td>Usage barrier</td>
<td>4.241</td>
<td>3.588</td>
<td>0.000</td>
</tr>
</tbody>
</table>

Table 3: Resistance levels across countries among non-users (latent scores).

<table>
<thead>
<tr>
<th></th>
<th>Finland</th>
<th>Portugal</th>
<th>Sig.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Overall resistance</td>
<td>4.516</td>
<td>3.907</td>
<td>0.000</td>
</tr>
<tr>
<td>Value barrier</td>
<td>3.420</td>
<td>3.083</td>
<td>0.000</td>
</tr>
<tr>
<td>Image barrier</td>
<td>2.583</td>
<td>2.287</td>
<td>0.000</td>
</tr>
<tr>
<td>Tradition barrier</td>
<td>3.514</td>
<td>3.828</td>
<td>0.000</td>
</tr>
<tr>
<td>Risk barrier</td>
<td>4.114</td>
<td>3.752</td>
<td>0.000</td>
</tr>
<tr>
<td>Usage barrier</td>
<td>4.727</td>
<td>4.011</td>
<td>0.000</td>
</tr>
</tbody>
</table>

In addition, differences between countries were computed only for non-users of mobile banking (Table 3). Out of the total number of 3597 responses 2490 respondents represented this group of
customers with 1075 and 1415 observations in Finland and Portugal respectively. The results of these responses follow the research results of the total sample. Therefore, the hypotheses H1, H2, H3, H5 and H6 are rejected and only the hypothesis H4 is supported by the data in terms of both the total sample and the sample of non-users.

6 CONCLUSIONS AND FUTURE RESEARCH

The structural equation model showed that usage and value barriers are the most intense determinants of overall resistance to mobile-banking, followed by image and risk barriers respectively. These results suggest that functional usability and relative advantage compared to other ways of banking are currently the most powerful inhibitors of mobile banking adoption. Interestingly, tradition appeared to be a negative determinant of resistance. Furthermore, Portuguese online bank customers showed less resistance in terms of usage, value, risk and image to adopting mobile banking services than their Finnish counterparts. However, Portuguese online bank customers showed greater preference for personal service, indicating more traditional banking behaviour compared to Finns. This idiosyncrasy of the Portuguese could mean a high pre-disposition to adopt new service channels alongside with more traditional ones.

Compared to Finland, the relatively low resistance scores to mobile banking among the Portuguese may reflect the fact that Portugal has simultaneously a low number of internet-connected computers and a high mobile penetration, a situation very auspicious for mobile services (Narinder, 2007). Another explanation for such surprising result might be related to the sampling method as only online banking users participated in the study. Rogers (2003) argues that adopter categorisation is based on innovativeness, i.e. the degree to which an individual is relatively earlier in adopting new ideas than other members of a social system. The fact that only the Innovators and Early Adopters of the total population in Portugal have so far adopted Internet banking, and that in Finland the diffusion of the innovation has already reached the Late Majority, may have resulted that, in general, the Portuguese sample consisted of more innovative individuals than the Finnish sample.

In general, innovativeness is related to demographics such as age. In our study the Portuguese sample consisted of much younger respondents compared to the Finnish sample. Future research is needed related to the role of innovativeness and demographic variables in mobile banking adoption. Moreover, Finland and Portugal represent very divergent countries in terms of cultural dimensions (e.g. Hofstede, 1980), hence providing good means to study the effect of culture (Kivijärvi et al., 2007). Future research could investigate the role of culture in consumer resistance to technological innovations.

REFERENCES


BUSINESS AND TECHNICAL WORKFLOWS FOR E-BUSINESS IN A VIRTUAL CLUSTER OF ISPS

Jane Hall and Klaus-Peter Eckert

Fraunhofer FOKUS, Kaiserin-Augusta-Allee 31, 10589 Berlin, Germany
jane.hall@fokus.fraunhofer.de, klaus-peter.eckert@fokus.fraunhofer.de

Keywords: Workflow technologies, e-Business, virtual organization, ISPs.

Abstract: Although workflow technologies have existed for some time, their true potential is only now being unlocked with the emergence of Web services technology and XML in conjunction with expanding Internet use. There has been extensive support for Web services and the standards have matured sufficiently for widespread adoption, replacing proprietary standards for interfaces and data that hindered the integration and automation of business processes, especially between partners wanting to cooperate in a virtual organization. This paper discusses the application of workflow technologies in supporting the e-Business of small Internet Service Providers (ISPs) collaborating in a virtual cluster. Not only business processes but also innovative technical processes are being executed as workflows in the marketing, deployment and operation of tailored ISP services. A scenario depicts the actual processes used and illustrates how the software being developed supports e-Business for ISPs collaborating in a virtual cluster.

1 INTRODUCTION

An ISP here is a small firm providing Internet and communication services to business users, themselves mainly small and medium enterprises (SMEs). e-Business is of direct concern to these ISPs for in the volatile and challenging marketplace of today it is imperative for them to improve their efficiency, to reduce their costs and to automate the selling and provisioning of the services they offer if they are to survive. For e-Business to be successful, full advantage must be taken of the technologies allowing open interoperability and supporting cross-organizational cooperation.

This paper discusses the use of workflow technologies to support a virtual cluster of small ISPs in marketing and delivering services to customers. The research has involved an investigation into how workflow technologies could be used in this context, whether this was a feasible objective, the challenges and issues involved, and how a workflow platform could be designed and implemented to support the objectives. Many of the issues are generic and applicable to many areas, they are however discussed in this paper within the framework of small ISPs collaborating in a virtual cluster. This paper illustrates some of the main ideas and issues in developing a workflow support system for ISP collaboration and in particular highlights an innovative part of the work by showing how technical processes that have been carried out manually by ISPs can be automated using workflow technologies. Only by using such automated processes can small ISPs hope to remain competitive and agile in the marketplace of today and the use of workflow technologies here can demonstrate their potential in such applications.

The structure of this paper is as follows. First some context is given to the rationale behind a virtual cluster of small ISPs. Then the application of workflow technologies to support the operation of the cluster and the collaboration between cluster partners in marketing and deploying services are discussed. The workflow technologies selected and the platform for modeling, specifying and executing them are presented. The business and technical workflows developed for running on the platform are examined. An e-Business scenario is outlined to show how the VISp software infrastructure is being used for marketing and deploying composite ISP services. The work is being carried out within the European IST project VISp (IST-FP6-027178) and conclusions are drawn on the experience gained in the project of utilizing workflow technologies for ISP e-Business.
2 RATIONALE FOR A VIRTUAL CLUSTER OF ISPS

The ISP market is changing rapidly, is quite volatile and very competitive. A small ISP lacks financial resources and has few personnel, which means it cannot cover all areas but does have expertise in specialized areas. It offers a limited range of services and so cannot cross-sell additional services to existing customers. No small ISP on its own can provide the geographic coverage and the wide range of specialized services that customers are now requesting.

Collaboration enables individual business entities to pool resources and competencies to provide value-added products and services (Camarinha-Matos, Afsarmanesh, 2006). The benefits of collaboration in virtual organizations generally are relevant also for small ISPs (Kürümlioglu et al., 2005). A solution for a small ISP is therefore to form a virtual cluster with other small ISPs for e-Business purposes. In this way, a greater variety of services can be offered as an ISP can integrate its own services with those of its partners in the cluster, extend its geographic coverage and become more visible in the market. As customers are requiring increasingly specialized services, often for a market of one, tailored services can be offered by the cluster to its customers by composing new services from the different services offered by various cluster partners.

The services offered by ISPs require a long-term relationship with the customer as they are provided over months or years. Unlike other forms of virtual organization (Camarinha et al., 2005), a VISP cluster is therefore intended to be a long-living entity, although partners can join and leave. Two operational modes are envisaged for the cluster. The Community mode is where each partner owns its own customers and the cluster is not visible externally. The partner serves its customers and the cluster is used dynamically as a pool of services for subcontracting. In the Virtual Enterprise mode it is the cluster that is visible to the customer and the cluster owns the customer relationship, the customer data and the customer transaction. When a customer makes an enquiry about a service in this mode it is sent to one partner, which communicates with the customer on behalf of the cluster. The cluster is a federation of independent partners collaborating in a decentralized manner to conduct e-Business. Each partner is a separate entity that joins in the collaboration to the extent that it wishes.

3 WORKFLOW TECHNOLOGIES

Workflow has been defined as “the automation of a business process” (Workflow Management Coalition, 1999). Ever since the existence of widely available computing power, there have been numerous developments aimed at automating and streamlining the activities in a business process in order to improve the efficiency of the organization, both internally and vis-à-vis other organizations (van der Aalst et al., 2003).

A major objective of the VISP project is to develop an innovative software platform enabling a cluster of small ISPs to collaborate and operate as a single business entity using workflow technologies to support the dynamic implementation and provisioning of tailored services. The provisioning and operation of ISP services by cooperating cluster partners requires the introduction not only of business but also of technical workflows in the cluster. However, little work has been undertaken on automating the activities in a technical process and it was one of the research topics to undertake work here so that technical processes could also be automated.

Workflow technologies are being adopted to provide the software infrastructure required by the cluster when undertaking e-Business transactions both with customers and between the partners themselves. Partners in the cluster can use collaboration-based process models to describe their cooperation. A workflow will specify how the individual roles participate in an end-to-end process. These processes are modeled as workflows using formal languages in order to be deployed and executed on distributed workflow engines.

There has been extensive support for Web services and the standards have matured sufficiently for widespread adoption, especially between partners wanting to cooperate in a value network (Keen and McDonald, 2000). Integration both within the enterprise as well as between enterprises is not only easier but also cheaper. These developments clearly have an impact on the VISP idea of cooperating roles in an SME cluster, which requires technologies that automate processes across organization boundaries. This availability is a significant element in making the VISP vision a reality.

Work that had been undertaken in the area of workflows, especially between organizations, could be used as a basis on for the work in VISP for the ISP service domain. This included projects such as CrossWork, which was concerned with cross-
organizational processes in the automotive industry (http://www.crosswork.info); CrossFlow, which investigated workflows in dynamically formed virtual organizations (http://www.crossflow.org); and Astro, which has been developing tools to support distributed processes during their lifecycle (http://astroproject.org).

A top-down approach is being adopted using workflow technologies. The high-level process can be recursively decomposed into sub-processes until the required level of detail is obtained for execution on a workflow engine. Various modeling languages were available and an investigation was undertaken to select those most appropriate for the VISP work (Eckert et al., 2006). BPMN was adopted as a modeling language for the high-level business and technical process modeling (BPMN, 2006). It is a standard containing a standardized mapping to BPEL4WS, or BPEL for short (Andrews et al., 2003), and it was felt that its graphical notation is more intuitive for non-IT specialists acquainted with the graphical notation of traditional business flowcharting notations. Despite the weaknesses and deficiencies of BPMN (Wohed et al., 2006), it was available on the market in tools complying with the standard at the time of language and tool selection for VISP. The ‘ideal’ solutions were not available and a pragmatic decision was to select a technology that seemed to have potential and was suitable for VISP’s purposes.

The BPMN flows are at a high level and may be decomposed to several levels of detail but they remain in BPMN. Mappings from BPMN to BPEL according to the standard are then undertaken and so tools are required that not only support the specification work in a particular language but that can map to another language and/or import and export such languages. BPEL also has its strengths and weaknesses (Wohed et al., 2003), but is in widespread use for implementing workflows. It was selected as it met the VISP requirements in the area of workflow technologies and also because the only standardized mapping was from BPMN to BPEL. The tools selected for these languages were also investigated and selected on the basis of their suitability for VISP aims in the first phase of the project as well as their adherence to the standard specifications of BPMN and BPEL.

In the last year both standards used in VISP have been significantly improved. OASIS published version 2 of WS-BPEL (BPEL, 2007) and several big IT companies proposed an initial submission of BPMN 2.0 (BPMN, 2008) to OMG. BPMN 2.0 will comprise standardized mappings to WS-BPEL as well as to the emerging standards for human interaction, BPEL4People and WS-HumanTask. Additionally the submission introduces a standardized XML-based exchange format for BPMN models, which was lacking in the previous version.

During the lifetime of the VISP project, most vendors of BPEL development and execution tools have migrated from BPEL 1.1 to WS-BPEL 2.0. Because the expressiveness of WS-BPEL and the number of built-in functions increased, it became an option to use the new version of the language and supporting tools. In this situation, the mapping chain has to be extended from BPMN to BPEL 1.1 and then further to WS-BPEL. Theoretically both mappings cannot be fully performed automatically. A mapping from BPMN to BPEL is only possible for a subset of BPMN models, thus not every valid BPMN model is “BPEL-valid”. Additionally not every BPEL process can be mapped to a WS-BPEL process. Fortunately the selected tool supports a best-effort mapping that works for most specifications and requires only minor manual improvements.

3.1 VISP Workflow Platform

As the VISP project is basing its software infrastructure on workflow technologies, the software platform that it is developing will allow the cluster to specify, model, deploy and execute workflows that support the operation of the cluster and the provision of tailored services to customers. This software platform consists of two major parts. The Workflow Modelling and Specification Platform (WiMSP) is designed to support service and workflow specification, modeling, choreography and orchestration. The Workflow Execution Platform (WiEP) executes and controls the workflows (see Figure 1).

Informal textual descriptions provided by domain experts and following Cockburn’s style of writing use cases (Cockburn, 2001) are the primary input to the WiMSP. They have to be formalized, first as BPMN models, then as abstract BPEL skeletons and emerge as an executable workflow that can be deployed on the WiEP. The workflow engines are the coordinating point of the WiEP that are responsible for executing and controlling the workflows specified by the WiMSP. The WiEP interfaces either directly or through mediation devices with partners’ ERPs and with network and system components.
3.2 VISP Workflows

Workflows are being developed in the VISP project to support the provisioning and operation of services offered by the partners in a VISP cluster. Processes are being specified, modelled, and executed as workflows of activities in both the business and the technical domains (see Figure 2).

A set of business workflows is being developed within the project able to deal with business processes in a dynamic cluster of partners and therefore providing for multilateral relationships and not only for binary relationships. These use standard processes where available, such as those from the OAGIS specifications together with their associated WSDLs (OAGIS, 2008) and are also based on the use cases produced as part of the requirements work undertaken within the project.

The technical processes in VISP are those that interact directly with network elements. Technical processes cover all technical activities related to the lifecycle of an ISP service in order to instantiate, commission, activate, deactivate and decommission the service. Further administrative activities such as testing, technical location transfer, suspend and resume are also being included. The project is providing formalized workflow specifications of technical processes that are currently manually executed in order to be able to process them automatically in a standardized way, something that has been possible for business processes but not so far for technical processes.

The technical processes are categorized in two groups, the Administrative Technical Processes (ATPs) and the Toolbox Technical Processes (TTPs). The ATPs perform all operations required to support a single ISP service instance lifecycle and to interface them to the corresponding business processes. They administer an instance of an ISP service and are invoked by the business processes. They can be defined and used without the VISP system and an adapter to the VISP system has been developed that maps and forwards all the necessary information stored in the VISP repositories and workflow engines to the ATPs and TTPs.

The TTPs act directly on low-level Web services found in the mediation servers of the various network elements. They usually contain just one atomic operation to be performed on one low-level Web service. They are invoked by the ATPs and are specific to a service instance type.

The relationship between the business processes, ATPs and TTPs can be seen in the service instance lifecycle where business processes are responsible for handling generic operations on composed services, or service sets, for example, “instantiate a service set”. Business processes first invoke the trading mechanism and determine the cluster partner that is best suited to provide the “Simple Call Service” VoIP. Then they invoke ATPs to perform operations on each of the service instances in the service set, in this case “instantiate Simple Call Service”. Each ATP can call directly one or more TTPs to carry out atomic administrative operations on service instances.
In this example, the ATP invokes the TTP that initializes the new user account in the network infrastructure. The TTPs help to isolate the upper layers of processes from any changes to the WSDL description as only the TTPs using the specified WSDL file need to be changed and not the ATPs themselves (see Figure 3). The parameters needed to invoke the TTPs are either stored in the VISP repositories or can be retrieved from the ISP’s employees, for example from network engineers, utilizing the VISP internal worklist management system.

The technical processes are intended to automate as much as possible of what network engineers have been undertaking manually to date. The need to be efficient and competitive means that small ISPs can no longer have their scarce human resources tied up in time-consuming tasks. The VISP project is therefore taking the technical knowledge of network engineers, preparing textual specifications of the steps involved in their work, modeling these steps in BPMN, mapping this to abstract BPEL, refining to executable BPEL and then running on the partners’ workflow engines.

Technical processes have not been standardized and so this work constitutes an innovative aspect of the project. The intention is to automate where possible not only the business but also the technical aspects of provisioning and delivering ISP services. This can save scarce human expertise for more complex tasks and thus enable ISPs to be more competitive and efficient in the marketplace.
4 REALISING E-BUSINESS WITH WORKFLOW TECHNOLOGIES

This section introduces the implementation of the VISP software infrastructure and the specific issues encountered. Implementation work is being carried out incrementally, with the first release concerned with how a simple VoIP service can be sold to a customer and how a VoIP system already accessible through Web services can be controlled from the VISP software infrastructure.

The e-Business scenario comprises the sale and provisioning of a VoIP service for a customer and it demonstrates how the various parts of the software infrastructure are utilized to support this. A prospective customer enquires about the ‘simple call’ VoIP service. A demonstration is provided to the customer, a simple VoIP service can be sold to a customer and it demonstrates how the various parts of the software infrastructure can be utilized to support this. A demonstration of a virtual service set is conducted with how a simple VoIP service can be sold to a customer. The demonstration includes how the various parts of the software infrastructure can be utilized to support this.

The sales representative uses the VISP Knowledge Base (SKB) for possible solutions and combinations of services. The SKB contains technical information about the services in the cluster and workflows have been developed to add, delete or modify a service as well as to browse the SKB. The Market Directory can also be searched to ascertain which partners are offering which services and the availability of these services. Once the component services required to support the VoIP service for the customer have been selected, the sales representative groups the services provided by cluster partners into a candidate service set. A request for service termination would typically imply three steps, deactivation, decommissioning and deinstantiation of the service instance so that its status becomes ‘historic’. Again, the technical workflows and deinstantiation of the service set are carried out automatically with the technical workflows developed in VISP.

If a cluster partner is not able to provide a service of the service set, an appropriate cluster member is selected by means of trading mechanisms performed according to a particular economic model. Trading is thus part of the service instantiation process and has a business (quote, contract) as well as a technical (instantiation) result. Details of the instantiated service set are stored in the Service Instance Base, which stores all trading and deployment information concerning the service instances comprising the service set. The technical part of the offer can be transferred to an ERP for a full offer preparation, if applicable. This can include all contractual terms and conditions, billing and payment details and SLA information.

The offer is then made to the customer and further negotiations may ensue. If the offer is accepted, the sales representative starts the commissioning of the service set. Commissioning of the service set has to be performed according to the requirements of the offer and in particular its timing. The resources previously reserved are allocated to the service set and once commissioning has been carried out, the service set can be activated on the date agreed with the customer. Commissioning and activation of the service set are carried out automatically with the technical workflows developed in VISP.

A request for service termination would typically imply three steps, deactivation, decommissioning and deinstantiation of the service instance so that its status becomes ‘historic’. Again, the technical workflows in the infrastructure carry this out.

An example of a centralized configuration of the VISP infrastructure for a two partner virtual enterprise mode is depicted in Figure 4. This configuration shows two partners A and B with their sales representatives and administrators. All actors are using partner-specific front ends that are connected to a common GUI server running the user interface to the VISP application that is written in Orbeon, an open source forms solution using XForms and Ajax technology. The Orbeon server can access the VISP repository services running as Web services on the repository server. Depending on the mode of the VISP cluster, either partner-specific repositories and/or global repositories can be queried. Additionally the Orbeon server invokes the business workflows running on an activeBPEL server as Web services.

These workflows invoke ATPs and TTPs running in the same engine via their WSDL interfaces. The last component shown in Figure 4 is the mediation server (MS_1) to the VoIP technical
infrastructure that can be accessed by VISP TTPs and by the ISP network operators via a proprietary administration interface (Jacinto). In operational environments, the repository servers and workflow engines will be replicated and distributed to the organizational domains of the cluster partners.

5 CONCLUSIONS

The research work of the VISP project has investigated the feasibility of using workflow technologies in an e-Business environment of ISPs. This paper has shown how workflows can support e-Business in a virtual cluster of small ISPs collaborating to offer and deliver services to their customers. A platform is being developed within the VISP project to model, specify and execute the business and technical workflows that have been realized. In assessing the research work, the following points can be made.

Experiences from the work in the project show that it is a challenge to extract the knowledge of experts in a form that is amenable to automation. The modeling and specification of textual specifications based on the knowledge of network engineers therefore has to undergo several reiterations.

Also, although there is a standardized mapping from BPMN to BPEL, the fact that BPMN is a graph-oriented language and BPEL a block-structured language has meant that conventions had to be established that were based on experience within the project as well as on work undertaken elsewhere (Ouyang, 2008). When using Web service technology it was also found necessary to agree on common conventions to ensure their interoperability and reusability. This comprises agreements on namespaces, message definitions, and binding styles. The WS-I Basic Profile was therefore adopted throughout the project (WSIO, 2004).

A big challenge in VISP is to interface generic business processes with dynamic technical processes. During the design time of the VISP application it is known that for each ISP service, corresponding ATPs with specific WSDLs have to be invoked. Unfortunately these WSDLs are unknown during design time. Thus a dynamic invocation mechanism has been developed that generates corresponding SOAP messages from information stored in the VISP repositories using XSLT transformations and retrieving endpoint addresses from the VISP UDDI. Such a dynamic invocation concept can be reused in every dynamic SOA environment.

The work in the VISP project is ongoing. Additional ISP services are being made available for offer and the trading models used are to be extended. More work is also required on ERP integration via OAGIS 9.x WSDL interfaces, and the configuration of the software needs to be based on a distributed architecture to replace the centralized one currently adopted.

Workflows are an interesting and innovative technology for supporting e-Business in the ISP domain and the marketing and provisioning of ISP services. Such technologies are being used increasingly for both intra- and inter-enterprise cooperation and ISPs, with their knowledge of Internet and Web technologies, are well-placed to use these technologies themselves. The research and development work within the VISP project has shown has shown that an infrastructure based on
workflow technology is feasible and can provide an effective automated environment allowing small ISPs to cooperate in offering and delivering services to customers. In addition, several technical solutions developed in the project can be transferred to SOA implementations in different application domains.

Despite the challenges encountered in implementing the workflow technologies for VISP objectives, an analysis of the project results to date show that workflow technologies have potential and can be used to support a virtual cluster of small ISPs. The VISP application is therefore being prototyped and will be further developed for use in a commercial environment.

ACKNOWLEDGEMENTS

This work was carried out with partial funding from the EU through the IST project VISP (IST-FP6-027178). Further information on VISP can be obtained from http://www.visp-project.org.

REFERENCES


Cockburn, A., 2001. Writing Effective Use Cases, Addison-Wesley, Boston, MA.


FACTORS AFFECTING THE USAGE OF T-GOVERNMENT SERVICES
An Exploratory Study

Michele Cornacchia, Filomena Papa, Stefano Livi1, Bartolomeo Sapio
Enrico Nicolò and Gaetano Bruno
Fondazione Ugo Bordoni, Via Baldassarre Castiglione, 59 - 00142, Rome, Italy
mcornacchia@fub.it, fpapa@fub.it, bart@fub.it, nic@fub.it, gbruno@fub.it
1University of Rome "La Sapienza", Faculty of Psychology 2, Via dei Marsi, 78 – 00185, Rome, Italy
stefano.livi@uniroma1.it

Keywords: Adoption of technology, interactive public services, digital terrestrial television (DTT), end-users evaluation, field study, home equipment, ICT influence on home life, predicting model, questionnaire, research methodology, residential users, T-government, usability, usage, use behaviour.

Abstract: Developing humanly acceptable information and interactive systems is a complex process where testing from the outset, the setting of operational criteria for efficiency, effectiveness and satisfaction, the development of prototyping etc., has to be considered as the central activity of a field implementation. Design and operative alternatives can be effectively selected and long-term targets can be seen to be attainable. However, sometimes it happens, especially in the domain of information system implementation, that obtained datasets can be effectively utilised to reach more tailored purposes. The case of interactive T-government services at the Municipality of Parma is analysed aiming to redesign the analysis model about the factors affecting the usage. To this respect one of the most recognized ICT predicting models in the literature is utilized. Final results show the success of the methodology and encourage further developments and applications in contiguous areas of ICT designed for residential population.

1 INTRODUCTION

The following study has been designed to be an exploratory study into aspects of T-government services usage among residential users. It aims to generate new groups of variables from a pre-existing dataset which can address the analysts towards different and reasonably more tailored purposes. The method and the first results here presented are part of a larger work still in progress, moreover having the objective to investigate the usability and economic factors affecting adoption and usage of T-government services.

With reference to the emerging ICT usage, it may happen, especially in the domain of information system implementation, that achieved datasets, even rich and widely aimed to support the design of technologies compatible with users’ abilities and needs, cannot be applicable to more complex and different kinds of analyses. Reasons can have different justifications, nevertheless it is usual that different theoretical positions do not all share the same views even in the same research domain. Hence there can be different orientations, even with respect to a similar target of an investigation, which can carry to a wide range of final resultant datasets, i.e. different for study typology, for structure, for procedures to submit/collect data, for user profile, for task design, for boundary conditions, and so on.

T-government stands for a wide set of services addressed to citizens (e.g. about health, education, tourism, bills payment), delivered by a Public Administration and accessible by Digital Television. The piloting research program on DTT (Digital Terrestrial Television) in Italy has been developed through a number of initiatives (Section 2.1), some of them issued to explore the potential impact on population of the new digital services associated with the broadcasted TV channels. In that field of investigation, T-government interactive services were introduced to be early tested with a real sample of residential private adopters of the digital TV decoder. A large amount of data were processed to investigate how residential end-users reacted to the usage of T-government services and they perceived their added value. Specifically, the results pointed
out how, in some cases, the usage of the services was to some extent low assessed. A significant difference in the usage was also found between informative and interactive services.

These findings have been the input for the current investigation about the factors affecting the usage. The study was carried out by comparing the UTAUT model (Unified Theory of Acceptance and Use of Technology, Venkatesh et al. 2003) with the real data of the basic questionnaire, the one applied to the Italian T-government project named “Parma Municipality Services” (Section 2.2), in order to redesign the analysis model.

The UTAUT model (Section 3) is a proved, reliable representation from social psychology, concerned with the determinants of acceptance and usage behaviour of new adopters of emerging information technology. It was applied as reference to cluster and re-code the relevant variables from the pre-existing dataset into new groups of indicators (Section 3.1). The indicators thus generated, were those of T-government, concerning respectively interactive services and informative contents. At last they were tested both for significance and construct power to check if they effectively meet the reference model basic requirements (Section 4), so verifying that structural differences may be sometimes overcome to start more complex analyses.

2 T-GOVERNMENT SERVICES

T-government services are herewith categorised as “informative services” (e.g. provision of general information) and “interactive services” (e.g. searching for specific information, e-mail, chat, form filling and form sending, payments).

Most of T-government services that have been so far developed are informative services. One of the opportunity given by T-government is to promote the use of ICT-based public services by large groups of people (e.g. the elderly), who have no Internet access or the required skills of using. For those people, the past experience with TV and remote control may be a key qualification to become effective users of the above services.

On the other hand some studies (Damodaran, 2002) make evident the gap of knowledge about the human aspects of T-government services: usage and usability, user satisfaction, the capability of services to reach all citizens and the capability to include them in the benefits of information society.

On these topics some field investigations have been developed in Italy in the framework of the T-government projects promoted by Centro Nazionale per l’Informatica nella Pubblica Amministrazione (CNIIPA) and Fondazione Ugo Bordoni (FUB).

2.1 T-government Field Investigations Developed in Italy

2.1.1 CNIPA Projects

In Italy, in 2004 CNIPA allocated 7 million euro in order to co-fund T-government projects based on DTT and addressed to the general public. As in the case of the advanced applications co-funded by FUB (3 million euro), such a competitive call came from a joint initiative of the Minister for Communications and the Minister for Innovation and Technology.

The experimentation carried out by CNIPA is mainly oriented to transfer on the DTT platform existing e-government services. Twenty-nine informative applications were selected and co-funded, with a total amount of investments exceeding 32.5 million euros. Correspondingly, 13 regions, 25 provinces, 164 municipalities and 15 mountain communities along with 40 local and national broadcasting corporations were involved. These funded projects aimed at the experimental implementation of several services for the citizens.

The services considered within projects can be grouped into five classes:

- social services: health care, social assistance and disability, third age, jobs and occupations;
- environment and tourism: environment, tourism, traffic congestion and cultural events;
- education: nursery schools, kindergartens, schools and universities;
- relationships with Public Administration: institutional activities;
- “Agorà”: citizens and elections.

Within several projects, interactive services exploiting back channel (reverse/return channel) and smart cards have been experimented.

As to the users involved, particular attention has been devoted to people having limited opportunities of accessing the Internet, users with disabilities, old people and people looking for their first job.

2.1.2 FUB Projects

One of the objectives of the six T-government projects co-funded by Fondazione Ugo Bordoni (FUB) was to experiment high interactivity T-government services, realised by Digital Terrestrial Television (DTT), involving real users.

The high interactivity T-government services provide the user with some of the following features:
remote interactivity using the return channel;
• high performance return channel (e.g. broadband or wireless);
• user authentication using a smart card (e.g. electronic identity card, services regional card or other kind of smart cards);
• on line payments.

The experimented services belong to different application areas: demographics, utilities and fines, education, T-health, T-learning, employment, T-commerce, T-banking.

In order to investigate usage, usability and socio-economical aspects of T-government services, an explorative field investigation was developed in each of the six projects.

The field investigations were realised using a common framework referring to human factors discipline (Papa & Spedaletti, 2001). The following main usability aspects related to interactive services were identified (Papa & Spedaletti, 2004): perceived usefulness, perceived ease of use and attractiveness, training and user support (human support, user manual, support provided by DTT, call center), user perception of technical disturbances and troubles (due to television signal, set top box, return channel), security and privacy perception (confidentiality of personal data, security of payments), impact of the equipment in the house, users’ satisfaction about the service including the comparison of different channels to perform the same task (e.g. DTT versus Internet, DTT versus traditional office desk).

Service usage is related to whether a given service is used or not and to the service utilisation level. Service usage is evaluated collecting subjective data provided by the user. The adopted indicators for service utilisation level are (Davis, 1993): frequency of use, time duration of the session, kind of use (shallow or intensive). Socio-economical aspects included: user profile (including income and social network information), TV and Internet usage (including other entertainment technologies), and scenarios (including interest levels, willingness to pay for equipment and services, decision factors).

The main results of the field investigations realised in the FUB projects can be summarised as follows:
• The role of the information and communication campaign is very important in ensuring a high adhesion to the field experiment by the users of the panels and, in some projects, in increasing the service usage.
• In general, the usage of T-government services has been quite low. A relevant difference in usage was found between informative services and interactive services. A possible explanation could be related also to the fact that some services with “high interactivity” are mainly referring to periodical or occasional activities (e.g. bill payment for phone etc, payment of fines, change of family doctor) which cannot justify a continuous use in the short time period of the field experiments (two-three months on average). However, the reasons of low usage should be investigated in depth.
• In general, technical problems were not producing relevant perception of disturbances and troubles and the users easily handled the provided equipment. Overall the users were quite satisfied with T-government services in comparison to different channels to perform the same task as the Internet and the traditional office desk. However, in some projects, technical problems produced relevant perception of disturbances and troubles mainly due to the lack of the digital television signal and to the interruption of the connection with the return channel. Other problems were connected with current input devices of the digital television, for instance user difficulties in the input of alphanumeric data by the remote control or by the virtual keyboard.
• Services developed in three of the six projects are still “on air” after the end of the project. These services have the common feature of being provided in the context of a public administration (municipality, regional administration, public schools).
• One of the initial objectives of the projects was to promote the use of Public Administration services by groups of people who have not the needed skills to use a computer and the Internet but are familiar with television and with the remote control. To this respect this objective was only partially reached since the investigation has shown that the most intensive users of T-government services were people using a PC and the Internet at home.

As mentioned before, the reasons of the low utilisation of interactive services need to be investigated more in depth. In particular, the identification of factors affecting usage of interactive T-government services could provide useful elements to increase their usage in the future.

2.1.3 The Field Investigation Developed in the Municipality of Parma

FUB provided the six projects with a common framework and some general tools:
• guidelines to evaluate usability of T-government services with high interactivity;
a first questionnaire for the users, about usability and usage aspects;
a second questionnaire for the users, concerning socio-economical aspects.

Each question or group of questions had the aim to evaluate indicators related to usability aspects, socio-economical aspects and usage. Within the above common framework, each project utilised different tools and techniques for data collection, adapting the above general tools to the particular context of the project. Each project was also responsible for data collection. In the following, the field experiment carried out in the Project “Parma Municipality Services” is described.

**T-government Services Implemented.** The project has developed both informative and interactive services. The former provide information about: Parma Municipality organisation; services offered to the citizens by the Municipality; cultural initiatives in progress in the city.

The interactive services allow the user (after authentication through Electronic Identity Card):
- to pay fines using a credit card;
- to visualise the state of a demographic dossier (e.g. for changing place of abode);
- to visualise the state of a request for education services provided by the municipality (e.g. nursery school).

A screen shot is presented in Figure 1.

**Figure 1: Data input to select a demographic dossier.**

During the field experiment a call center was available to the users for any information and help in the utilisation of the services.

**The Users Panel.** A sample of 200 people was selected from a group of 4,000 citizens of Parma Municipality. The selection followed two main criteria:
- demographic characteristics to reflect the main features of citizens in the Municipality of Parma (in terms of age, gender, and city district);
- potential interest toward the interactive services under experimentation (holders of driving licence and electronic identity card).

It was decided to choose the sample age ranging from 20 to 45 years, people in general skilled in the use of new information technology. The 200 citizens were randomly selected in this age range. Finally, 181 citizens (88 males and 93 females) took part at the experiment for the all planned period of two months and a half.

**Procedure.** In the first phase, the T-government applications were implemented and broadcasted in the territory of Parma Municipality.

In the second phase a randomly chosen users panel of 200 citizens was selected with the above mentioned criteria. The recruitment officially started when the municipality sent a letter to the users’ panel explaining objectives and modalities of the trial. The users were then invited to contact the call center to choose between self-installation and personnel assisted installation of the decoder.

In case of self-installation, users were asked to pick up by themselves the Set-Top-Box (STB) at the Municipality offices. A user guide was also released to aid home installation and services use. In case of assisted installation, a technician brought the STB to the people residence and there trained the users.

**Tools and Techniques for Data Collection.** Data were collected mainly using a paper-based questionnaire filled in by the users. The questionnaire was both delivered and collected by Municipality of Parma personnel.

**Results.** The first results of the field investigation can be summarised in the following points:
- The usage of interactive services has been quite low. In particular only 20% of the users answering to the questionnaire completely used the interactive services. 35% used the interactive services in a shallow way. A possible explanation could be related to the fact that the services are referring to occasional activities (e.g. payment of fines).
- Technical problems were not producing relevant perception of disturbances and troubles and the users easily handled complex equipment. Overall the users were quite satisfied with T-government services in comparison to different channels to perform the same task, as the Internet and traditional office desk. However, the sample involved in the field study was composed by a
large part of young people using ICTs. Of course it is not possible to generalise such indications to the whole Parma population and, even less, to the whole Italian population.

3 THE FACTORS OF USAGE

Since many years the study/assessment of usage of emerging ICTs requires to be carried out through the analysis of more and more reliable models (Taylor & Todd, 1995).

3.1 Theoretical Background

In this paper we assume that the main forecasting feature of the most recognized models in the literature, i.e. the “Technology Acceptance Models” (TAM) (Malhotra & Galletta, 1999) and UTAUT (Venkatesh et al., 2003) in particular, could be applied to analyse other experimental datasets, built up by different criteria other than, of course, to be suitable with ICT contexts of use. This hypothesis, whether verified, should expand the possibilities to analyse the usage constructs of the original sets of variables and as well as to come to more satisfying results. In fact, it is very difficult to attain data due to the high complexity and variability especially about socio-technical systems (Dillon, 2000) and/or when technology is being oversold (Cornacchia, 2003). The relevant literature on this topic describes the development of several models of technology acceptance by the users and many extensions to the basic constructs (Malhotra & Galletta, 1999; Venkatesh & Davis, 2000), mostly built with the behavioural elements (Ajzen, 1996) of who is forming an intention to act, and then free to act without limitations. UTAUT extends TAM by introducing the terms of social influence and cognitive instrumental processes and, above all, it ultimately unifies the main competing user acceptance models, namely eight theoretical approaches sharing the same basic concepts. As shown in Figure 2, the key dependent variable is the intention and/or the use of technology, thus the final purpose of the model is that of understanding and providing explanation of the phenomenon “use of information technology”, taken as dependent variable. The role of intention, as an antecedent of behaviour (i.e. of use), is the actual critical factor of the model and was analysed and proved in literature.

![Figure 2: Basic concept underlying user acceptance models.](image)

In the formulation of the unified model four constructs were identified which may play a significant role as direct determinants of the user acceptance and usage behaviour: Performance Expectancy, Effort Expectancy, Social Influence and Facilitating Conditions. In addition, Age, Gender, Experience and Voluntariness of Use were the moderators of the model.

3.2 Recoding T-government Variables

In the current exploratory study all direct determinants of UTAUT constructs were considered in the application with experimental T-government dataset. To this extent, the questionnaire items were presented with new technological devices, a number of factors influence their decision about how and when they will use them, notably two main key sets of constructs: Perceived Usefulness and Perceived Ease of Use (Davis, 1989).

Perceived Usefulness is defined as the user’s “subjective probability that using a specific application system will increase his/her job performance within an organisational context”. Perceived Ease of Use refers to “the degree to which the users expects the target system to be free of efforts”. Both complex constructs predict the attitude toward using the system, in that order defined as “the users desirability of using the system”.

There can be also strong behavioural elements with which the models assume that when someone is forming an intention to act, then he/she is free to act without limitations. UTAUT extends TAM by introducing the terms of social influence and cognitive instrumental processes and, above all, it ultimately unifies the main competing user acceptance models, namely eight theoretical approaches sharing the same basic concepts. As shown in Figure 2, the key dependent variable is the intention and/or the use of technology, thus the final purpose of the model is that of understanding and providing explanation of the phenomenon “use of information technology”, taken as dependent variable. The role of intention, as an antecedent of behaviour (i.e. of use), is the actual critical factor of the model and was analysed and proved in literature.
recoded into aspects dealing with usage of “STB”, “informative contents”, “interactive contents” and others. During the matching and re-coding process, the following four step scheme was applied:

1. set up the reference sets of complex constructs from UTAUT;
2. search for simple items in the questionnaire that can replicate the causal structure of the unified model UTAUT, according to its requirements and basic conditions;
3. re-code by grouping them into new sets;
4. test statistically both for significance and meaningfulness of the new constructs of variables.

Following steps 2 and 3 in the above mentioned scheme, factors of UTAUT model were replicated using the indicators adequately selected from the T-government questionnaire. The result of such recoding process is herewith summarised.

**Use Behaviour**
- Use of Parma municipality services
- Use of DTV services
- Frequency of use
- Time duration of each session
- Type of use (intensive/shallow)
- Comparison with alternative channels
- Weekly interactions

**Behavioural intention**
- Willingness to pay for equipment
- Willingness to pay for services
- Interest levels toward DTV services

**Performance expectancy**
- Perceived service usefulness
- User satisfaction
- Comparison with alternative channels
- Factors influencing adoption
- Privacy and security

**Effort expectancy**
- Perceived ease of use
- Attractiveness
- Perception of technical problems
  - Loss of television signal
  - Too much time for application loading
  - Loss of connection by the return channel
  - Problems in establishing the connection
  - Too much time to reach the objective
  - Technical troubles in using equipment
- Interaction with equipment
- Service functions
- Impact of equipment in the house

**Social influence**
- Social network

**Facilitating conditions**
- Decoder installation
- User support provided by DTT
- User manual for STB install

- Manual for service use
- Information on the cost of connection
- Call center and contact center

**Moderators**
- Gender
- Age
- Income and education
- Experience (skills in using TV, Internet and entertainment appliances)

Arranged by the above recoded factors, since this was an exploratory study, dependent variables considered in testing the significance of the new causal constructs (step 4 of the procedure) were:
- \( \text{cont}_\text{utente_servizi} \) (STB Use)
- \( \text{cont}_\text{interattivi} \) (Interactive Service Use)
- \( \text{cont}_\text{informativi} \) (Informative Service Use).

Besides, in carrying out assessment, following independent variables constructs were applied:
- Performance expectancy
- Effort expectancy
- Social influence
- Facilitating conditions

The other constructs left over, i.e. Behavioural Intention and Moderators, were not taken into consideration at this preliminary stage of the study. Finally, ahead of computing each predictor of the novel UTAUT model, the three dependent variables were standardised in order to make them comparable each other.

### 4 RESULTS

In order to better understand the active relations among the new generated set of variables, either for the general or for the information or for the interactive usage of T-government services, a regression analysis was applied according with the UTAUT reference model. This choice was suitable given the low variability affecting the new dependent variables. To this regard, it was pointed out how people didn’t use many of the several available on-line facilities, but they typically used no more than one at time.

#### 4.1 Predictors Estimation

Regression analysis allowed to estimate the different weights of the UTAUT model factors applied to the novel T-government usage structure, that is the influence each of them may have in controlling the weight of a single predictor.

Three models were thus arranged and tested, using as criteria the assessment respectively of STB
Use, of Interactive Service Use, and of Informative Service Use. As shown in Table 1, set cont_utente_servizi (STB Use) as dependent variable, all coefficients with the exception of Facilitating conditions were significant (overall model: Adjusted R Square=.13; p<.000). Hence, it can be said, the more subjects perceived high values of these three predictors in adopting the STB the more positive was the evaluation of this dependent variable (Beta=.279, .311; .236; overall p<.001).

Table 1: Dependent variable: general STB Use.

<table>
<thead>
<tr>
<th>Predictors</th>
<th>Beta</th>
<th>t</th>
<th>Sig</th>
</tr>
</thead>
<tbody>
<tr>
<td>Performance expectancy</td>
<td>.279</td>
<td>3.406</td>
<td>.001</td>
</tr>
<tr>
<td>Effort expectancy</td>
<td>.311</td>
<td>3.793</td>
<td>.000</td>
</tr>
<tr>
<td>Social influence</td>
<td>.236</td>
<td>3.368</td>
<td>.001</td>
</tr>
<tr>
<td>Facilitating conditions</td>
<td>.046</td>
<td>.648</td>
<td>.518</td>
</tr>
</tbody>
</table>

In the second model of Table 2, the evaluation of Interactive Service Use, set cont_interattivi as dependent variable, gave nearly the same results than the previous one (overall model: Adjusted R Square=.14; p<.000). Therefore, still excluding Facilitating conditions, the more subjects perceived high values of the three predictors in using the Interactive Service the more positive was the evaluation of this dependent variable (Beta=.315, .285; .264; overall p<.001).

Table 2: Dependent variable: Interactive Service Use.

<table>
<thead>
<tr>
<th>Predictors</th>
<th>Beta</th>
<th>t</th>
<th>Sig</th>
</tr>
</thead>
<tbody>
<tr>
<td>Performance expectancy</td>
<td>.315</td>
<td>3.900</td>
<td>.000</td>
</tr>
<tr>
<td>Effort expectancy</td>
<td>.285</td>
<td>3.528</td>
<td>.001</td>
</tr>
<tr>
<td>Social influence</td>
<td>.264</td>
<td>3.811</td>
<td>.000</td>
</tr>
<tr>
<td>Facilitating conditions</td>
<td>.032</td>
<td>.456</td>
<td>.649</td>
</tr>
</tbody>
</table>

Finally, as shown in Table 3., the evaluation of Informative Service Use, set cont_informativi as dependent variable, found that no one predictor turned out to be effective (overall model: Adjusted R Square=.003; p=n.s.); (Beta=.107, .153; .081; .024; overall p>.05).

Table 3: Dependent variable: Informative Service Use.

<table>
<thead>
<tr>
<th>Predictors</th>
<th>Beta</th>
<th>t</th>
<th>Sig</th>
</tr>
</thead>
<tbody>
<tr>
<td>Performance expectancy</td>
<td>.107</td>
<td>1.222</td>
<td>.223</td>
</tr>
<tr>
<td>Effort expectancy</td>
<td>.153</td>
<td>1.749</td>
<td>.082</td>
</tr>
<tr>
<td>Social influence</td>
<td>.081</td>
<td>1.086</td>
<td>.279</td>
</tr>
<tr>
<td>Facilitating conditions</td>
<td>.024</td>
<td>.315</td>
<td>.753</td>
</tr>
</tbody>
</table>

According to this last model, Use of Informative Services is not a variable depending on the given UTAUT factors. Whether a predictor should be taken for a supplementary analysis, the best rank is Effort Expectancy.

4.2 The Final Model

By means of regression analysis it was demonstrated if and how much the values taken by the dependent variables, as well arranged from the T-government dataset, either depended on or were determined by the correspondent values taken by the independent variables, in our case the main determinants of UTAUT. Figure 3 summarises the results of the positive dependences found in the analysis.

The figure also shows that Facilitating Conditions hasn’t effective links to the hypothesised usage constructs.

5 CONCLUSIONS

In this paper the main feature of a forecasting model has been applied in order to give a further representation of a field research dataset about ICT information and interactive services proposed for residential users. Specifically, data coming from an Italian T-government field experience at the Municipality of Parma were re-processed to bring to more significant results towards new purposes, i.e. the identification of most relevant factors affecting usage. In such a way, the low usage of interactive services scored at Parma, and as well as of informative services and decoder adoption, was explained in terms of “use of information technology” key dependant variable, in UTAUT mentioned as direct determinant of user acceptance and usage behaviour. The determinants Performance Expectancy, Effort Expectancy, Social Influence and Facilitating Conditions were reformulated by means of T-government data to assess if and how much the information system was humanly acceptable developed. Results demonstrate that the procedural scheme can be successfully applied to recode most basic questionnaire items into new constructs.
referable to UTAUT. The resultant model achieved is statistically significant for the first three constructs above mentioned, while it reveals to be partially adequate for the last construct. Results encourage also to go ahead in further developments on some important issues.

At this stage of the study about the usage of T-government services, the UTAUT key moderators (i.e., age, gender, experience, voluntariness) were not included in the process. However, a further analysis has been planned to aim at finding whether it is possible to incorporate in the model the variables “income” and “education”, either as new moderators or as new determinants.

This study is antecedent to a development also in the area of forecasting the T-government services diffusion processes for residential users. Actually, re-modulated datasets and UTAUT constructs can effectively feed a microsimulation model (Turk et al., 2008), where different policy measures (such as subsidisation or inclusion of different services) could be tested from different aspects, for instance their cost, time and relevance.

ACKNOWLEDGEMENTS

The T-government projects promoted by FUB and CNIPA were co-funded via a competitive call coming from a joint initiative of the Italian Ministry of Communication and the Italian Ministry for Innovation and Technology. The field investigation of the Project “Parma Municipality Services” was developed in the framework of the agreement between Fondazione Ugo Bordoni and the project partners, Municipality of Parma, Itcity and Enterprise Digital Architects. The authors would like to thank: Sebastiano Trigila (Chief of the FUB T-government program) for the supervision of the projects; Roberto Azzano (ANFoV) and Giulia Berni (ANFoV) for their contribution to the supervision of user trials.

REFERENCES


Cornacchia, M., 2003. Usability, a way to distinguish from the good, the bad, and the irrelevant in the web, Cost 269, Conference Proceedings, Media Centre Lume, University of Art and Design, Helsinki (Finland), pp.159-163, September.


BUSINESS PROCESSES MANAGEMENT USING PROCESS ALGEBRA AND RELATIONAL DATABASE MODEL

Kelly Rosa Braghetto, João Eduardo Ferreira
Department of Computer Science, University of São Paulo, Rua do Matão 1010, 05508-090 São Paulo, Brazil
kellyrb@ime.usp.br, jef@ime.usp.br
Calton Pu
College of Computing, Georgia Institute of Technology, Atlantic Drive 801, 30332-0280 Atlanta, U.S.A.
calton@cc.gatech.edu

Keywords: Business Process Management, Workflows, Process Algebra, Relational Databases.

Abstract: Integrating information systems with tools that manage workflows and business processes is not always a simple task. This difficulty becomes more accentuated when the execution control assumes countless business processes. This work presents NavigationPlanTool (NPTool), a tool to control the execution of business processes that can be easily integrated into the information systems. NPTool is supported by Navigation Plan Definition Language (NPDL), a language for business processes specification that uses process algebra as formal foundation. NPTool implements the NPDL language as a SQL extension and offers two other important services: processes instantiation and process instances execution monitor. This paper describes the NPTool showing how the process algebra features combined with a relational database model can be used to provide a scalable and reliable control in the execution of business processes.

1 INTRODUCTION

Business Processes Management (BPM) involves methods, techniques and tools to support the project, the execution, the management and the operational analysis of business processes (Leymann et al., 2002). While the traditional definitions of workflow place emphasis on the execution of operational processes, BPM also gives support to the diagnosis phase, allowing the processes to be analyzed in order to detect flaws and possible improvements to the project.

In this context, associating formal frameworks with the process project phase is valuable since they provide non-ambiguous models, improve the diagnosis capability and enable a reliable execution control of the process.

Although there are a number of tools based on formal frameworks directed to the management of workflows and business processes, integrating these tools with other applications isn’t always easy. Some of these tools are exclusively developed for modeling and simulation of processes, as occurs in CPN/Tools (Beaudouin-lafon et al., 2000). In other cases, the tools effectively carry out the execution control of business processes, but they do not provide mechanisms that allow them to be easily used within the information systems that manage business processes.

This work presents the NavigationPlanTool (NPTool), a tool that offers mechanisms for the representation and execution control of business processes supported by a process algebra formalism (Fokkink, 2000). NPTool uses the Navigation Plan Definition Language (NPDL) (Braghetto et al., 2007) and a relational database to specify business processes and to control their instantiations and executions. NPTool implements NPDL as an extension of SQL; this implementation allows an easy integration with traditional information systems, which generally already have mechanisms that facilitate the access to RDBMSs. The storage of the processes data in a relational database adds scalability to the execution control provided by NPTool. Moreover, the processes definitions can be reused in different applications. In this context, the database maintained by NPTool can be viewed as a common repository of processes.

Section 2 presents a summary of the related works and the reasoning behind the use of NPDL. Section 3 describes the services offered by the NavigationPlanTool; in particular, Section 3.1 describes the
data structures in the relational database model to represent the processes, and Section 3.2 explains the mechanism for execution control. Using a real application as example, Section 4 illustrates the execution control applied to an instance of a process. Finally, Section 5 discusses the contributions of this work.

2 RELATED WORK

Management technologies for business processes were developed to meet the following needs: (1) to promote the separation between the specification of the process and its implementation; and (2) to take away from the applications the responsibility of the execution control of business processes. Among the various existing languages to define business processes, it is not easy to see a formal pattern as a representation basis, one that is capable of expressing in a non-ambiguous way the semantics associated to the existing constructions in these languages. The formalisms regarded as natural candidates for this role are the Petri nets and the process algebras. Works such as (Aalst, 1998), (Puhlmann and Weske, 2005) describe the use of these formalisms in the specification of business processes.

The Yet Another Workflow Language (YAWL) (Aalst and Hofstede, 2005) is the most well succeeded approach for business process specification based on Petri nets. It was developed intending to provide means for defining all the original control-flow patterns described in (Aalst et al., 2003). Also with the intention of representing business processes with a formal basis, Navigation Plan Definition Language (NPDL) (Braghetto et al., 2007) was created. NPDL is based on the concept of navigation plan of RiverFish architecture (Ferreira et al., 2005) as well as in the operators of ACP (Fokkink, 2000), a member of the family of process algebras. The navigation plan was formalized in (Ferreira et al., 2006) as a set of all business processes demanded from an application in order to achieve business goals. Like in processes algebras, processes in NPDL are defined by algebraic expressions. The expression of a process is built based on NPDL operators and steps (atomic actions or processes); the operators indicate the execution order of the steps. For completeness, we will provide an overview of NPDL operators.

2.1 NPDL Operators

The basic control-flow patterns described in (Aalst et al., 2003) can be easily represented with three basic operators of NPDL:

- **Sequential Composition** ("*"), the process term \( A.B \) means that the activity \( B \) will be enabled for execution after the completion of the activity \( A \). An activity can be an atomic action or a subprocess;
- **Alternative Composition** ("+"), the process term \( A+B \) means that initially both activities \( A \) and \( B \) will be enabled for execution, but only one of them can be executed;
- **Parallel Composition** ("||"), the process term \( A|B \) means that the activities \( A \) and \( B \) can be executed parallelly.

These basic operators were directly extracted from ACP (Algebra of Communicating Processes) and have their semantics formally defined by a set of transition rules that will be presented in the Section 3.2. In order to represent all control-flow patterns identified by Aalst et al in (Aalst et al., 2003), the NPDL was improved with additional operators:

- **Interleaved Parallel Composition** ("||+"), the process term \( A||+B \) means that the activities \( A \) and \( B \) can be executed in any order (e.g., \( A.B+B.A \)), but not in parallel;
- **Multi Merge Composition** ("&&"), the process term \( A&&B \) means that the activity \( B \) will be enabled for execution after the completion of each thread of control of activity \( A \);
- **Discriminator Composition** ("&&\text{true}\"), the process term \( A&&\text{true}B \) means that the activity \( B \) will be enabled for execution after the completion of the first thread of control of activity \( A \);
- **Unlimited Repetition** ("*n"), the process term \( A*n \) means that the activity \( A \) can be executed an unrestricted number of times;
- **Number Limited Repetition** ("n\#", where \( n \) is a positive integer number), the process term \( A\#5 \) means that the activity \( A \) must be executed five times;
- **Function Limited Repetition** ("f\#", where \( f \) is a function that returns a positive integer number), the process term \( A?f_{1} \) means that the activity \( A \) must be executed the number of times calculated by function \( f_{1} \) at execution time;
- **Conditional Execution** ("%r\#", where \( r \) is a rule, e.g. a boolean function), the process term \( %r_{1}A \) means that the activity \( A \) will be enabled for execution if the return value of the rule \( r_{1} \) is true at execution time;
- **Negative Conditional Execution** ("%!r\#", where \( r \) is a rule, e.g. a boolean function), the process term \( %!r_{1}A \) means that the activity \( A \) will be enabled for execution if the return value of the rule \( r_{1} \) is false at execution time.

Providing more details about NPDL is beyond the scope of this paper; the definition of the language and the specification in NPDL of each one of the 20
control-flow patterns described in (Aalst et al., 2003) can be seen in (Braghetto et al., 2007).

3 NavigationPlanTool

NavigationPlanTool (NPTool) provides methods for storing actions and processes in a relational database and for controlling the instantiation and execution of these processes. The tool offers operations like creation/removal of instances and services for monitoring the navigation plan execution. These services are also responsible for storing logs of the execution of navigation plans in the database and for recovering executions that have been interrupted before completion.

The programming language used in the implementation of NPTool was Java (Java 2 Platform Standard Edition - J2SE 5.0). NPTool extends JDBC API - Java DataBase Connectivity Application Programming Interface. JDBC enables Java programs to execute SQL commands and to interact with databases that are compatible with SQL standard. The usage of JDBC turns NPTool into a RDBMS independent tool. Since it was developed as a library of functions, the NPTool can be easily integrated into other Java applications.

NPTool is composed of three services:

- **NPDL Interpreter**: receives an input command and makes the lexical, syntactic and semantical analysis. There are two possible situations: (1) the command is a NPDL valid command, and (2) the command is not a NPDL valid command. In the former case, the interpreter will translate the command to pure SQL commands before the submission to the RDBMS. In the latter case, the command will be directly passed to the RDBMS. The translated SQL commands will be executed over a relational database environment whose tables are created by the interpreter to store processes, actions and instances data;

- **Process Instantiation Service**: provides functions for creating process instances. A process instance represents a request to a specific process. All the instance data, as well as process definition data associated with the instance, are stored in a database;

- **Process Instance Execution Monitor**: is responsible for linking a process instance to its execution data (navigation plan). This service contains the functions that control the execution of the navigation plan of a process instance.

3.1 The Relational Data Structures Created by NPDL Interpreter

The relational data model does not have appropriate structures for representing processes and, therefore, it requires an additional data structure to achieve this task.

In (Braghetto et al., 2007), the NPDL syntax was defined as an extension of the SQL syntax. The main goal of implementing NPDL as a SQL extension was to enable a RDBMS to create and handle business processes, providing to information systems an easy access to these features. The data structure kept stored in a database by NPTool is represented in the extended entity-relationship diagram shown in Figure 1. The diagram shows the data structures needed for representing business processes and for controlling their instantiations and executions.

In NPDL, the navigation plan of a process is defined by an algebraic expression formed by steps composed by operators that indicate the execution order of these steps. The relationship-set NAVIGATION_PLAN_REL between the entities-set PROCESS and STEP represents this definition. The specialization of STEP is total and disjoint, i.e., each entity in STEP is a process, or an action, or an operator, or a rule, or a function or a number. As described in Section 2.1, rules are always associated with the operators “%” and “%!”.

3.2 The Execution Control of Business Processes in NPTool

When an execution of a process instance is started, the navigation plan associated with the instance is recovered from database. According to the algebraic
expression that represents the navigation plan, an expression tree of the instance is built. Expression trees are used in the execution monitor service to determine the execution order of the steps in the navigation plan. In this work, the expression tree of a process instance was called navigation tree. A navigation tree node can represent one of these three elements: a NPDL operator, an action or a process. A navigation tree is a complete binary tree; its internal nodes represent binary operators of NPDL, whereas its leaf nodes represent actions or processes. As the navigation tree nodes represent the possible execution steps, an important attribute of a node is its current status. A node, in a specific point of the instance execution, can be in one of the following states: not started (N), started (S), finished (F) or canceled (C). The Figure 2 shows an hypothetical example of one of the possible navigation tree for an instance of process $P = (a + b).c.(d||e).f$.

After creating the navigation tree, the execution monitor recovers the current state of the instance execution by consulting the execution instances log, that is stored in the database. This operation only is done for instances that have already started their execution, but have not finished it yet. When an instance execution is started, all nodes in its navigation tree will have the current status not started. If the instance execution have already been started in a previous execution of the monitor service, there will be records related to the instance in the log and they must be loaded to the navigation tree. This is made by updating the status of the tree nodes associated with the steps indicated as started, finished or canceled in the log. Each step performed in the navigation plan of an instance results in the insertion or update of a record in the log.

The "navigation" through the tree of an instance determines the execution order of the steps of this instance. The algorithms of the monitor service of NPTool use the semantics of NPDL operators and the transition rules of process algebra to visit a navigation tree. Figure 3(1,2,3) shows the transition rules extracted from (Fokkink, 2000) that define the operational semantics of terms of basic process algebra and ACP. The operational semantics of a language describes how a valid sentence must be interpreted in sequential steps. Figures 4 to 6 show how the transition rules are applied on the navigation tree branches to guarantee that the instance status, after starting or finishing the execution of an action, will be consistent with the behavior specified by the algebraic expression of the process associated with the instance.

The four rules in Figure 3(1) states that the process defined by the term $x + y$ terminates successfully after executing $x$ or $y$. Figure 4 shows how the alter-
native composition operator ("+") is treated in a navigation tree branch. As in figures 7 to 14, the nodes delimited by an ellipse with dashed line indicate the actions that are enabled for execution in the current instance state represented by the tree. The labeled arrows signalize the start or the end of the execution of an action\(^1\). So, \(S(a)\) signals the start of the action \(a\), while \(F(a)\) signals the end of the same action. The start or finish of the execution of an action in an instance modifies the status of a leaf node in its navigation tree. Each modification in the status of a leaf node in a navigation tree requires the update of status of its predecessor nodes. To enable this operation, each node keeps a pointer to its parent node, besides the pointers to the left child and the right child nodes. The status node configuration in a navigation tree determines the current state of the instance; thus, a change in the instance state generates a new set of actions currently enabled for execution. The status of a node that represents an alternative composition operator is specified by the status of its child nodes in the following way:

- **Started**: if the left or the right child has status started;
- **Finished**: if the left or the right child has status finished;
- **Canceled**: if the left and the right child has status canceled.

The process term \(x||y\) indicates that the terms \(x\) and \(y\) will be executed parallelly, i.e., it is possible to execute an initial transition of \(x\) (\(x \rightarrow \sqrt{1}\) or \(x \rightarrow x'\)) or a initial transition of \(y\). This behavior is formally specified by the rules in Figure 3(3). The set of all the possible states originated by an expression involving the parallel composition ("||") operator is large, as Figure 6 shows. The status of a node that represents the parallel composition operator is specified by the status of its child nodes in the following way:

- **Started**: if the left or the right child has status started;
- **Finished**: if the left and the right child has status finished;
- **Canceled**: if the left and/or the right child has status canceled.

The conditional execution, negative conditional execution, unlimited repetition, number limited repetition and function limited repetition operators are not treated as nodes in the navigation tree. They are treated as attributes of nodes, since they influence the execution of the complete branch rooted at the node with which they are associated. If a node have a rule or a function associated with it, then it is necessary to execute this rule or function before “visiting” the node in the navigation algorithm. The execution of the rule or function, like in the case of atomic actions, is a responsibility of the applications that use the NPTool. When an action or a process term is delimited by some of the two conditional execution operators, the rule that conditionates the execution is attributed to the node that represents the delimited action or to the root node of the delimited process term.

In the example of Figure 7(a), the execution of action \(a\) is conditioned by the return value true of rule \(r_1\), while the execution of \(b\) is conditioned by the return value false. In this example, the return value of \(r_1\) was true; so, after finishing the execution of \(r_1\), only

\(^1\)In NPTool, the actions of a process may not have an instantaneous execution, i.e., the execution can be distinguished by a start event and by an ending event.
the action $a$ was enabled to be executed.

In the case of repetitive actions or process terms, the node that represents the action or term is labeled as repetitive, and if this repetition is limited by a number or a function, this number or function will be associated with the node. Figure 7(b) shows how a repetitive node is treated by "expanding" it in an equivalent branch.

The inspection of a navigation tree in order to get enabled steps to execution in an instance is also controlled by the transition rules expressed in Figure 3; the rules are used to determine how the nodes of the tree should be visited. The execution monitor starts to visit the navigation tree from its root node. Accordingly the current status of the tree nodes, the navigation returns a set of steps currently enabled to be executed in the instance.

Only the basic operators of NPDL are kept in the navigation tree. The other operators are treated by a mapping algorithm that removes them of the navigation tree, replacing them by an equivalent branch containing only basic operators. Figures 8 and 9 show how nodes rooted at non-basic NPDL operators in the navigation tree can be replaced by other equivalent
4 EXAMPLE OF EXECUTION

This section illustrates a simulated execution of a process instance. It shows how works the navigation that gets the enabled steps and updates the state of the navigation tree. The process used as example is a simplified version of a system developed for controlling the acquisition of items to a library collection.

Consider the set of rules and the set of actions belonging to the acquisition system that are defined, respectively, as \( R = \{ r_1 \} \) and \( A = \{ a_1, a_2, a_3, a_4, a_5 \} \), such that:

- \( a_2 \) gathers the prices of the order items;
- \( a_3 \) liberates the order for purchasing;
- \( a_4 \) registers the purchase receiving;
- \( a_5 \) registers a problem with the purchase.

The acquisition system can be specified by the process \( P \) defined in NPDL by the commands:

\[
\begin{align*}
\text{SET } P_1 & = a_1 \cdot a_2 + \{ a_1 \cdot a_2 \} \cdot P_1; \\
\text{SET } P & = \& r_1 P_1 \cdot a_3 \cdot \{ a_4 + a_5 \} . a_4; \\
\end{align*}
\]

The check step of order data is treated as a rule (in the case, \( r_1 \)). Thus, the execution of the activities involved in the acquisition system will only be enabled if the order data is valid and there is available budget for the purchase. The subprocess \( P_1 \) encapsulates the actions of sorting items (\( a_1 \)) and gathering items price (\( a_2 \)). The actions are composed by the operator “\(|\ast|\)”, that indicates that they can be executed in any order, but not in parallel. Moreover, \( P_1 \) was defined in a recursive way. This indicates that these actions can be repeated an unrestricted number of times before the liberation of the order for purchasing. After the liberation of the order (action \( a_3 \)), it is possible that problems occur in the purchase. These problems are registered in the acquisition process by the action \( a_5 \). The repetition operator “\(|\ast|\)” was used associated with \( a_5 \), to represent the possibility of occurring more than one problem during the purchase.

The initial navigation tree of an instance of acquisition process is represented in Figure 10(a). Each node has, at the upper left side, a character that indicates one of the following states: “N” (not started), “S” (started) and “F” (finished). It is important to notice that, associated with the node that represents the process \( P_1 \) in the tree, exists the rule \( r_1 \). Such as the tree nodes, a rule has a status too. Associated with the node representing action \( a_5 \), exists the operator “\(|\ast|\)”, that indicates that the action can have its execution repeated.

In the beginning of the execution, the tree is visited starting on its root node (distinguished by a dashed line in Figure 10(a)). While the navigation algorithm is visiting the tree nodes, it comes across the rule \( r_1 \) and it adds the rule in the set of currently enabled steps to execution. Since the execution of the other steps depends on the execution of step conditioned by \( r_1 \), no other step can be executed at first.

Continuing the visitation to the tree nodes, the al-
algorithm detects the existence of the operator “|∗” in two different places in the tree (nodes x and y in Figure 11(a)). Since the operator “|∗” is not an NPDL basic operator, it requires a special treatment.

In NPDL, a process term like $a|∗b$ is equivalent to $a.b+b.a$. Figure 11(b) shows the navigation tree after mapping the operators “|∗”. After the mapping, the visitation algorithm is able to identify the currently enabled steps to execution. As the distinguished nodes in Figure 11(b) shows, there are two enabled steps to execution in the navigation tree: actions $a_1$ and $a_2$.

Consider that, after the execution of $r_1$, the action $a_2$ have been executed, and the algorithm for propagating the status update through the tree nodes is applied on the nodes $x$ and $z$ of Figure 11(b). In the resulted tree, only the action $a_1$ is enabled to execution; Figure 12(a) shows the tree after the execution of $a_1$.

In a navigation tree, branches rooted at a node whose status have already been defined as finished and which do not have enabled steps to execution can be removed from the tree. The goal of this removal is the optimization of the main memory occupied by the storage of the trees. Figure 12(b) illustrates the tree resulted by the removal of the inaccessible branches of the tree in Figure 12(a); the nodes distinguished by the background color gray are the nodes that had their children removed.

The navigation for getting the currently enabled steps to execution detects the node $y$ in Figure 12(b), which once more represents the process $P_1$. The node is replaced by the navigation tree of process $P_1$, resulting the tree in Figure 13(a). In the tree in Figure 13(a), the nodes $x$ and $y$ represents the operators “|∗”, which must be mapped to basic operators, resulting the tree in Figure 13(b). In the tree in Figure 13(b), the enabled steps are the actions $a_1$ (nodes $w$ and $y$), $a_2$ (nodes $x$ and $z$) and $a_3$ (node $v$). If the chosen step for execution is the action $a_3$, the application of the algorithm for propagating the status update through the tree nodes over the node $v$ in Figure 13(b) results in the tree in Figure 14(a).

When the algorithm of navigation finds the node $y$ in the tree in Figure 14(a), it will apply the mapping of the operator “?∗” associated with the action $a_5$, resulting the tree in Figure 14(b). In this tree, the currently enabled steps to execution are the actions $a_4$ and $a_5$. The tree in Figure 14(c) is obtained by executing the action $a_4$ and applying algorithm for propagating the status update over the node $w$ in Figure 14(b). The root node of the tree in Figure 14(c) has the status finished, which indicates that the execution of the process instance was terminated successfully.
NavigationPlanTool is a framework that uses NPDL and a relational database model to specify business processes and control their instantiations and executions. The NPDL language is based on the concept of navigation plan of the RiverFish architecture as well as in the operators of processes algebra.

The simplicity of the relational database model (used by NPTool for the storage of processes, instances and its executions) is consequence of the adoption of processes algebra as formal basis, due to its algebraic and textual form. Albeit simple, this model has proved to be quite flexible, since it allows the enrichment of the expressiveness of NPDL (by including new operators in the language) without causing impact to the data structure. This model also allows us to make reuse of the processes definition between different applications that use the NPTool, as it can be viewed as a common repository of processes. Moreover, the compositional characteristic of processes algebra makes the composition of great processes from smaller ones possible. This constitutes an important aspect for business processes, in which the reuse of definitions is a common occurrence.

The implementation of NPDL as an extension of the SQL language provided by NavigationPlanTool allows its easy integration to traditional information systems, which already have mechanisms that facilitate the access to RDBMSs. In addition to that, the storage of the processes definition data, instances and its executions in a relational database makes the NavigationPlanTool more robust and scalable.
Figure 13: Navigation tree (a) after substituting \( P_1 \) and, next, (b) after the substitution of the operators \("|*"\).

(a) (b)

Figure 14: Navigation tree (a) after executing action \( a_3 \), (b) after the removal of inaccessible branches and mapping operator "?*"; next, (c) after executing action \( a_4 \).

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)

\( a \)
cation and reuse of control-flow patterns in business processes, data flow management for business processes and automated generation of NPDL expressions from graphical representations.

REFERENCES


ON THE USE OF “QUALIFIED” DIGITAL SIGNATURES

Maurizio Adriano Strangio
Department of Mathematics, University of Rome “Roma Tre”, Italy
strangio@mat.uniroma3.it

Keywords: Qualified Digital Signatures, Key Agreement Protocols.

Abstract: The European Commission Directive 1999/99/EC aims to provide a community-wide framework for the use of electronic signatures and thus for promoting electronic trade and communication among the member states. The directive introduces the notion of “qualified” digital certificates as a means to maintain legal effects of digital data that are analogous to those of paper-based signatures. To this end, proofs of (physical) identity and possession (of the private key) are fundamental requirements that must be fulfilled by the requester during the public key enrollment process.

Digital signatures are often employed as secure building blocks in key agreement protocols that use public key authentication. The need for the rigorous analysis of such protocols has recently emerged; there are currently several formal models of distributed computing that may serve for this purpose. However, we point out these models employ rather trivial or unpractical approaches in the modeling of the procedures and policies employed by certification authorities.

We believe that usage of qualified certificates not only should represent the standard practice for CAs in order to sustain secure electronic commerce (and in general all forms of secure communication) but also represents the first step towards the domain of a global PKI.

1 INTRODUCTION

Symmetric key algorithms have a troublesome shortcoming — the need to generate and distribute a secret key to the parties who wish to privately communicate. Furthermore, keys must be exchanged prior to any communication by making use of a (out-of-band) secret channel and not via a public network.

With public key cryptography, in a network of $n$ principals, each party needs only to distribute a public key to her $n-1$ peers so the total number of keys to manage is $O(n)$, rather than $O(n^2)$.

However, with asymmetric key algorithms public keys can be easily exposed (since they are delivered in unencrypted form) therefore a principal using the wrong key may be totally unaware that a malicious party is posing as her intended peer. As the result, public keys must be certified in a separate process prior to any communication, possibly by a trusted third party; for the parties willing to communicate this is logically equivalent to having an authenticated channel available to distribute the public keys in order to receive assurance of each others purported identity.

In practice, this service is provided by a Public Key Infrastructure (PKI) which is an arrangement that entails a trusted third party (a.k.a. certification authority — CA) fulfilling a notary function by binding the public key (and a set of informative attributes) to the owner of the corresponding private key. The outcome of the process is a digital certificate (e.g. X.509v3), signed by the authority itself, that provides undisputable evidence of the true identity of the principal holding the private key/public key pair.

While this approach is technically straightforward, the management (e.g. operation and funding) of a PKI can become quite intricate (see Ellison and Schneier, 2000) for an overview of the difficulties that may be encountered) so that the public key validation process may be almost as difficult as distributing secret keys (after all symmetric key cryptography has been profitably employed since the era of Caesar).

Whether a CA may be regarded as trustworthy (which ultimately is a decision of the relying party) is a major concern and is not just a matter of considering certain distinguishing elements such as the reputation or the nationality of the CA. It turns out that there are other requirements that must be satisfied and often these are not clearly stated in the CA
policies (e.g. in the Certification Practice Statements — CPS) or, even worse, not even completely understood. For this reason, in some countries (e.g. Italy) there are government agencies entitled to perform the final accreditation of CAs (which are also subjected to periodic inspections to verify full compliance with the law).

In fact, the main thesis of this paper is that a “physical proof-of-identity” and “proof-of-possession” are both mandatory to maintain the legal status of digital signatures. With a physical proof-of-identity the certificate requester must provide corroborative evidence of her identity (e.g. an id-card) in a “face-to-face” registration procedure at the registration authority (RA). With a proof-of-possession the subscriber must prove he exerts (exclusive) control over the private key corresponding to the public key undergoing the certification.

Although the “physical” identification process is often perceived as an unduly burden by most subscribers (although it is essentially identical to applying for a driving license) it is beneficial for a number of reasons:

- it discourages from making use of PKIs for criminal activity;
- facilitates law enforcement agencies in the prosecution of the crimes related to signature/encryption schemes (e.g. identity theft, fraud);
- is also valuable for assessing the trustworthiness of individuals.

On the one hand, the proof-of-possession requirement has been the subject of many discussions (see (Asokan et al., 2003; Lauter and Mityagin, 2006)) with the majority of PKI standards firmly encouraging its use (observe that when the CA generates the public/private key pairs this requirement is immediately fulfilled). On the other hand, many CAs will issue certificates offering different levels of identity assurance; a physical proof-of-identity is required for the certificates that are used in applications that need to establish the purported identity of the subscriber with certainty (e.g. commercial transactions); this is opposed to certificates issued with email addresses used as identification credentials (which are suitable for non critical applications).

However, often an explicit indication of how the subscriber was identified is omitted from the certificate attributes thus opening subtle vulnerabilities in its use (the relying party may unwillingly omit to perform any validity checks).

We stress that a proof-of-possession merely implies that the applicant has access to the private key corresponding to the target public key; therefore a physical proof-of-identity is essential for the legal recognition of digital signatures.

CA policies often refer to the PKCS#10 standard (PKCS#10v1.7, 2000) which describes a syntax for certification requests; an entity applying for a certificate may be prompted by the RA to digitally sign an electronic “certificate request” that will be subsequently conveyed to the CA for the final certificate handout. Such an arrangement subsumes a proof-of-possession. Surprisingly, the standard also includes the provision for requests using paper forms; this implies that CAs should define alternative means in their policies for the requester to prove knowledge of the private key.

2 DIGITAL CERTIFICATES IN THE EUROPEAN COMMUNITY


In particular, the legal recognition of electronic signatures entails the use of advanced electronic signatures based on “qualified certificates”, which are created by secure (signature-creation) devices (e.g. tamper-proof standard-compliant devices such as smart cards holding the private key), to ensure they:

(a) satisfy the same legal requirements of handwritten signatures on paper-based data (refer to the definition of advanced electronic signature in the Directive);
(b) are admissible as evidence in a court of law.

The majority of European member countries have already embodied the directive in their legal systems (civil laws).

Qualified certificates are certificates that comply with the rules contained in Annexes I and II of the aforementioned Directive and appear to satisfy both the proof-of-possession (Annex I, letter (e)) and proof-of-identity (Annex I, letter (c)) requirements as we discussed in the preceding section.

Under such arrangements we call the resulting certificates “strong qualified certificates”; as opposed to "weak qualified certificates” which may be supplied by CAs that do not strictly comply with both the above requirements (e.g., consider a CA performing a proof-of-identity by means of an on-line zero-knowledge identification protocol).
3 KEY ESTABLISHMENT PROTOCOLS THAT USE DIGITAL SIGNATURES

Digital signature schemes (and other public key encryption schemes) are often employed as secure building blocks in key establishment protocols that use public key authentication for the purpose of ensuring the parties involved in the communication that the session key was established with the intended peer and not with an impostor. Once a session key is available, the data exchanged is authenticated and undisclosed against third parties thus allowing secure communications over an insecure network.

Traditionally, cryptographic protocol security has been a matter of perceived confidence supported by heuristic proof arguments and by the protocol surviving many years of public scrutiny. More recently, the approach has changed and the priority has switched to developing formal proof frameworks. In this context, a main line of research employs complexity-theoretic models for distributed network computing and is dedicated to key establishment protocols (Bellare and Rogaway, 1993; Blake-Wilson and Menezes, 1998; Shoup, 1999; Bellare et al., 2000; Canetti and Krawczyk, 2001; LaMacchia et al., 2006; Diffie et al., 1992). Such models employ both private and public key cryptographic techniques.

In the above models most of the fine level details of PKIs are abstracted away; while in principle this is a reasonable approach (we already mentioned that PKIs are an intricate subject on their own) we often see that important topics such as the key registration procedures and policies employed by CAs are disregarded by many authors. We believe that such issues are not only tightly related to the correct operation of the protocol but may also eventually lead to (legal) disputes among the participants and therefore must be appropriately settled.

As a starting point for our discussion, let us recall three public key registration procedures commonly encountered in the literature (LaMacchia et al., 2006):

a. Honest key registration. All parties (including those controlled by the adversary) follow the key generation procedures honestly and register the resulting public keys before engaging in any communications. The adversary can corrupt parties only after key registration has completed;

b. Proof-of-possession. An authority performs some validity check upon public key registration. In particular, a party is required to prove knowledge of the corresponding secret key. The adversary can register public keys for corrupted parties at any time;

c. Arbitrary key registration. Parties can register arbitrary public keys (even the same key as some other party) without any validity checks. The adversary can register public key for corrupted parties at any time.

We comment that case c. is unrealistic since it is extremely unlikely that a CA will ever accept to enroll public keys for which the requestor has not provided any validity check; the minimum requirement is verifying the identity of the applicant.

Item a. exemplifies the behavior one would normally expect from honest principals holding valid certificates. The case of interest here is the compromise of the private key of a principal whether or not he is aware of this fact (the adversary may be able to subtly obtain a copy of the private key).

The proof-of-possession requirement invoked by item b. is not sufficient to achieve adequate levels of security (as discussed above); we have already pointed out that without a proof-of-identity CAs will not issue certificates. We now elaborate further on this point.

Lauter and Mityagin (Lauter and Mityagin, 2006) have recently presented protocol KEA+ that, as opposed to the original version (KEA), is resilient to unknown key share (UKS) attacks; as a countermeasure, they suggest to include the identities of the protocol participants as arguments of the key derivation function (KDF) and also claim that the above countermeasure avoids the need for a proof-of-possession.

Recall that a UKS attack involves a (man-in-the-middle) adversary, posing as a legitimate party (say A) in a protocol run between honest parties (say A and B), that is able to convince one party (e.g. A) to accept her identity (C) while the peer (B) is unaware of this fact (i.e. B thinks he’s interacting with A). This vulnerability is regarded of interest whether or not the adversary is able to have the attacked party accept a chosen session key.

We believe there are two issues that are overlooked in their arguments. Firstly, the adversary must provide a valid proof-of-identity; therefore, unless she is willing to reveal her true identity (otherwise she must be able to perfectly disguise herself and also offer a counterfeit id card — indeed a non trivial task), she may be liable of being legally pursued for her actions in a court of law. Observe also that the (physical) proof-of-identity policy rules out the well known online UKS attack described by Kaliski (Kaliski, 2001) against the MQV protocol (Law et al., 2003) (notice that in this attack the adversary is able to have the target party accept a chosen session key).

Secondly, the lack of a proof-of-possession (albeit
not being necessary for a protocol to resist UKS attacks) implies loss of the non-repudiability property for digital signatures; at a later time the signer can eventually deny having participated in the protocol run (and therefore having established a particular session key with a peer).

We may conclude that the KEA+ protocol (and many others alike) are not suitable for electronic trade and commerce and at best can be used within the realm of a corporate domain for the exchange of data between employees (the proof-of-possession would be satisfied anyway since the keys are generated and managed in house).

From the above remarks we see that the primary sources of concern derive from the legal implications involved in electronic data processing facilities (which constitute the underlying principles of the European Directive).

The lessons learned are twofold:

- strong qualified certificates are necessary to prevent legal disputes;
- if legality is not a concern peers may use weak qualified certificates.

Therefore, relying parties must be able to verify certificate policies of the CA to determine whether the certificates are suitable and trustworthy for a particular application.

As a consequence, CAs should also have clearly stated policies for appropriately setting certificate attributes; for example, in strong qualified certificates the X.509v3 Key Usage extension (RFC3280, 2000) should always specify only one use for the public key (namely for signature validation) and have the non-repudiation bit asserted.

We mention the recent attempt of Boldyreva et al. (Boldyreva et al., 2007) to set out rigorous models of the public key enrollment process with a CA. In particular, they show that a simple challenge-response interactive protocol suffices to fulfill the proof-of-possession requirement. Notice that a fundamental assumption in their work is the existence of an authentic channel between the user and the CA; although (as the authors admit) it would be otherwise difficult to establish any reasonable security claim this hypothesis enormously reduces the practical usefulness of the registration protocols (e.g. running it on an open network environment is risky — similar arguments apply to the proof-of-possesson protocols found in the technical literature for the Internet (RFC2875, 2000)).

As a possible enhancement to the registration protocols, one may consider the CA assigning a unique transaction key TK to the subscriber when the physical proof-of-identity is performed; the TK may be used only once either as a shared session key (provided it is cryptographically strong enough — say, at least 128 bits) or as an access token to be included as an argument of the registration protocol on the user side (and verified by the CA) thus providing an authenticated communication channel.

4 CONCLUSIONS

In this paper we have (re)affirmed the need for CAs to require mandatory proofs of identity and possession (of the private key) from subscribers requesting digital certificates for public keys in order to preserve the legal status of the associated signature schemes.

We have also reviewed the notion of qualified certificates as introduced by the European community directive 1999/99/EC (European-Parliament, 1999). As far as electronic documents are concerned (e.g. contracts) the directive says that all member states should ensure that advanced electronic signatures (i.e. signatures that are linked to qualified certificates) must be given the same legal effect as paper-based signatures and are thus form valid evidence in court trials.

We have also pointed out that formal models of distributed computing, where the security of key establishment protocols using public key authentication is evaluated, often adopt a rather trivial approach when considering public key enrollment procedures.

Our main thesis is that large-scale CAs (offering national or international cross-border services) should only offer qualified certificates (with the appropriate proofs established) whether they are relative to digital signatures used to sign documents or employed as building blocks in key establishment or other cryptographic protocols. Other types of certificates, which we have broadly classified as weak qualified certificates, only grant a limited degree of security (assurance) with no legal effects and can be eventually used in particular domains (e.g. the PKI used in a home-banking system).

Although (in principle) one could choose the appropriate certificate depending on the application (among those offered by a CA) there would be significant benefits if all the players involved could assume that certificates delivered an equivalent degree of assurance rather than having to rely on non-standard policies. To this end, it would be desirable that strong qualified certificates be adopted as a universal standard for digital signatures, thus delivering further impulsion to the development of interoperable world-wide PKIs.
ACKNOWLEDGEMENTS

The author is grateful to the reviewers for their helpful comments and suggestions.

REFERENCES


RESEARCHING SEARCH

A Study into Search Engine Marketing Practices in Ireland

Chris Barry and Debbie Charleton

Department of Accountancy and Finance, National University of Ireland, Galway, University Road, Galway, Ireland

chris.barry@nuigalway.ie, d.charleton@nuigalway.ie

Keywords: Information search, Search Engines, Search Engine Marketing, Website optimization, ethics.

Abstract: Researchers have identified the Web as the searchers first port of call for locating information. Search Engine Marketing (SEM) strategies have been noted as a key area for firms to consider when developing, maintaining and managing their Websites. SEM falls into two categories - ‘Search Engine Optimisation’ (SEO) and ‘Paid Search’. To examine how these SEM methods are reflected in practice, and the issues inherent in carrying out SEM effectively, a survey was conducted amongst small to medium enterprises (SMEs) in Ireland. The research reveals that Irish SMEs plan to spend more time and resources on SEM in the future. Most firms utilize an informal SEM strategy, where SEO is perceived to be the more effective technique in driving traffic to a company Website. Respondents cite the use of ‘keywords in title and description tags’ as the most used SEM technique; while ‘Pay for Placement’ was found to be the most widely used Paid Search technique. In concurrence with the literature, measuring SEM performance remains a significant challenge with many firms unsure if they measure it effectively. An encouraging finding from the study is that Irish SMEs adopt a positive ethical posture when undertaking SEM.

1 INTRODUCTION

For many, online search has become an important daily activity to navigate the Internet and has triggered fundamental changes in the way people seek and find information (Browne, Pitts and Weatherbe, 2007). Boughton (2005) and Fain and Pedersen (2006) have identified the Web as the searchers first port of call for locating information. Search Engines regularly feature amongst the most frequented Websites and are a vital mechanism by which e-commerce sites promote themselves (Jansen and Molina, 2006). Given the importance of the Search Engine as a method for finding information, and the opportunity it presents to advertisers as a marketing medium, Search Engine Marketing (SEM) strategies have been noted as a key area for firms to consider when developing, maintaining and managing their Websites (Sen, 2005).

2 STUDY BACKGROUND

SEM has emerged as the main method firms use to successfully increase the visibility of their Website. SEM consists of a number of methods that broadly fall into two categories - ‘Search Engine Optimisation’ (SEO) and ‘Paid Search’ (Boughton, 2005; Jansen and Resnick, 2005; Feng, Bhargava, and Pennock, 2007). SEO is the process of identifying and fine-tuning elements of a Website in order to achieve the highest possible visibility when a Search Engine responds to a relevant query (Xing and Lin, 2004; Zhang and Dimitroff, 2005). SEO leverages the Search Engine algorithm to channel users to specific Websites, the results of which are referred to as organic results. Paid Search uses Search Engines to trigger a display of advertisements based on the topic or keyword entered (Laffey, 2007).

A major factor in choosing one strategy over another is the searchers’ perception of sponsored listings (Paid Search) over that of organic or editorial listings (SEO). Jansen and Resnick (2005) suggest that searchers tend to trust the relevance of organic links over sponsored links, whereas Boughton (2005) argues that searchers more commonly following sponsored advertisements. Sen (2005) uses an analytical model to assess the use of Paid Search over SEO. He finds that even if the cost of Paid Search and SEO were the same, firms would still select Paid Search as the most popular strategy.
In contrast, Xing and Lin (2004) find SEO more popular based on cost and due to searchers’ perceptions of Paid Search results. Jansen and Molina (2006) surmise that a dual strategy of SEO and Paid Search is the best approach having found that Paid Search results are just as relevant as non-paid results (SEO). Laffey (2007) discusses how a strategy may depend on which Search Engine is subscribed to. He reports that Google users are more likely to click on organic results and least likely to click on paid results, whereas MSN users are more likely to click on Paid Search results.

Several optimizing SEO techniques have been identified. Some of these include: providing keyword rich Website content throughout all Web pages (Seda, 2004; Schultz and Fristedt, 2005; Dawson and Hamilton, 2006); using keywords in the title and description meta tags in the Website code (Zhang and Dimitroff, 2005); and keeping the likes of flash, graphics, forms and frames to a minimum (Schultz and Fristedt, 2005; Dawson and Hamilton, 2006). Mentz and Whiteside (2003), Oliva (2004), Seda (2004), and Schultz and Fristedt (2005) recommend a link development strategy whereby other good quality and relevant Websites are used to develop or implement a link back to the company Website.

Paid Search takes a number of different forms. Paid Placement is where, in response to a particular search term, a firm pays a fee for their link to appear on the search results page, for example Google’s AdWords. Paid Inclusion is where Search Engines are paid to index pages from a Website (Thurow, 2007). These results can often be mixed into the organic/editorial listings on the search engine results page (SERP). For example, Yahoo and other Search Engines firms offer Paid Inclusion programmes for this purpose. At present Google does not offer such a service (Mangalindan, 2003). With Paid Submission, advertisers pay Search Engines to speed up the process of having their Website reviewed more quickly than they would under a free submission (Sen, 2005). However there is no guarantee that it is accepted and included in the SERP (Thurow, 2007). Contextual search is another type of Paid Search whereby search results are returned based on user behaviour and relevance rather than matching keywords (Vine, 2004), for example Google’s AdSense programme. Local search allows advertisers to target local customers in a geographic region rather than marketing to a generic audience (Jones, 2006). Google permits this feature through its AdWords programme.

There is widespread recognition of the importance of using performance metrics for SEO and Paid Search in order to achieve a competitive advantage (Weischelbed, Mateard and Deans, 2005; Fain and Pedersen, 2006; Laffey, 2007). However despite these calls there is general consensus throughout the literature regarding the mis-use of measurement tools for SEM (Weischelbed et al, 2005). Some SEO measures put forward are: setting a baseline to measure inbound links to a site, where a company should continue to improve the number and quality of inbound links to increase the site’s page rank; and site usage statistics (Schultz and Fristedt, 2005) where the likes of Google Analytics can be used. Zhang and Dimitroff (2005) and Seda (2004) recommend tweaking different SEO aspects of the Website and then measuring the change in rank of the Website on the SERP on a continual basis. Paid Search measures put forward include: Cost Per Click (CPC) where the cost to an advertiser to generate one click is monitored; Cost Per Mille (CPM) where the cost to display an advertisement a thousand times is recorded; Cost Per Action (CPA) which monitors the cost for actions such as signing up for a new account or making a sale; and Conversion Rate which is the ratio between CPC and CPA (Fain and Pedersen, 2006). Smith (2002) discusses conversion metrics as a key tool for measuring ROI. Another important and commonly used measure is Click Through Rate (CTR). CTR is the ratio between the number of times a Web link is displayed against the CPC (Kumar and Shah, 2004; Fain and Pedersen, 2006).

Ethical practice in SEM is an increasingly important consideration for all stakeholders (Palmer, 2005), particularly searchers who rely on Search Engine to organise and distribute returned results in an ethical way (Zimmer, 2006). Ethics takes two forms - the responsibility of Search Engine operators and the onus on Search Engine marketers to conduct themselves ethically. Conflictingly, it is in the best interest of Search Engine operators to push for paid advertisement (that generate revenue) rather than organic results. This leads searchers to be concerned about SERP results since some Search Engines do not disclose the difference between sponsored and organic search results (Moxley, Blake and Maze, 2004) - a clear breach of any assumed ethicality. Other concerns are expressed about: unethical link development such as when a site is found with identical link development on other sites that are completely unrelated; bogus blogs generated to include links to increase the Website’s rankings (Economist, 2006); ‘Black Hat SEO’ where popular
links are rented out from quality Websites and hidden links are placed on prestigious sites unknown to their owners; and Web spamming where invisible or unreadable text, or excessive use of the same word are used to fool the Search Engine spider into thinking there are certain keywords on the Website (Esparza, 2007). Click fraud is also discussed at length in the literature (Asdemir and Yaha, 2006; Laffey, 2007). It occurs when an advertiser falsely generates clicks on an ad with the only goal of increasing the payment to the advertiser.

3 RESEARCH METHODOLOGY

The broad objective of this research effort was to examine SEM strategies of small to medium sized enterprises (SMEs) in Ireland. It was also designed to: identify the specific SEM techniques used; reveal which strategies are perceived as being more effective; discover if SMEs use performance metrics; to investigate if firms consider an ethical perspective when undertaking SEM activity; and to identify the main challenges faced in undertaking SEM. The research method used was an extensive, quantitative survey. The population from which the sample was drawn possessed the following attributes: an SME operating in Ireland with an in-house marketing manager or specialist in a marketing department and an operating company Website. The sampling frame was selected based on the research questions, as well as the attributes listed above for selection. The sample information was compiled from Kompass, a business database listing companies in Ireland. A total of 95 responses were received out of a sample of 623, giving a response rate of 15.24%. The data collection method chosen for this study was a self-administered postal questionnaire and a matching Web-based questionnaire.

4 FINDINGS

4.1 Website Development

Most firms that participated in the survey develop and manage their Website internally. While one third outsource the development of the Website they still manage and update their Website in-house. The significance of this is that most respondents have the capacity to carefully manage their SEM strategy with expertise based in-house. For these firms technical aspects of SEM should not be a barrier to understanding or developing online marketing techniques.

4.2 Nature of the SEM Strategy

Concurring with recommendations elsewhere (Ostler, 2001; Porter, 2002), two thirds of those surveyed describe their SEM strategy as closely aligned with their firms overall business strategy. This finding suggests firms have moved beyond experimentation and approach a more mature relationship between their online and broader strategies.

Many respondents (46%) revealed their SEM strategy was of an informal nature, while (39%) reported a fairly formal or a formal SEM strategy. The remainder indicated that they did not carry out any SEM strategy. This contrasts with the exhortations of writers that recommend a strategy formulation to achieve sustainable competitive advantage (Varadarajan and Jayachandran, 1999; Novak, Hoffman and Yung, 2000; Schultz and Fristedt, 2005; Sen, 2005). In relation to the effectiveness of their SEM strategy, nearly three quarters of respondents agreed they would consider their SEM strategy ineffective if it did not return their company’s Website link on the first page of the SERP. This finding is reinforced by research that suggests searchers view at most the first few pages of the SERP (Lempel and Moran, 2000; Zwick, Rapport, Lo and Muthukrishnan, 2003).

An overwhelming number (91%) listed Google as the Search Engine most subscribed to for indexing by companies. This is followed by Yahoo at 65%, with MSN coming significantly behind (29%), while AOL, Alta Vista and Ask only received a very small percent. These findings are in line with findings elsewhere by Schultz and Fristedt (2005) and Taylor (2007). Schultz and Fristedt noted that Google receives the highest number of keyword searches followed by Yahoo. The findings here suggest Irish firms are aware of the importance of subscribing to the most popular search engines to target the largest possible market. It may also suggest that many of these SME’s are subscribing to Google because it is free, whereas subscription to other search engines such as Yahoo and MSN incur cost.
4.3 Types of SEM Strategy Used by SMEs

The literature suggests a key factor in a firm’s choice of SEM has much to do with how searchers perceive organic versus sponsored results (Seda, 2004; Boughton, 2005; Jansen and Resnick, 2005; Sen, 2005). If firms are reading searchers correctly, they are choosing to focus on SEO. Nearly half of respondents indicated that they predominantly use SEO for SEM (see Figure 1). This concurs with a European report conducted by SEMPO (2007a), which found that French, Spanish and Italian advertisers focus more on the use of SEO over Paid Search. In a similar type report on the US market, SEMPO found that over three quarters of respondents use SEO as their main form of SEM (SEMPO, 2007b). A third of those surveyed here report that they use a combination of SEO and Paid Search, while few respondents reported that they use Paid Search only. This latter finding would suggest a clear view of Irish SMEs that Paid Search on its own is not capable of delivering searchers to their Website. The position would seem to be vindicated by a recent Amarach Consulting report showed a clear majority (82%) of searchers admitting to rarely or never clicking on sponsored results when conducting a Web search (Amarach, 2007).

4.4 SEO Strategy Techniques

Respondents cite the use of ‘keywords in title and description tags’ as the most used SEO technique, followed by use of ‘keywords throughout the whole Website’ (see Figure 2). The first of these findings contrasts with the recommendations of Seda (2004) and Schultz and Fristedt (2005) who state that keywords should be used throughout all pages of the Website content. While Dawson and Hamilton (2006) advise minimizing the use of graphics, forms and frames, a significant number (22%) employ this technique. The use of these elements often restricts spiders from reading the Website, thereby affecting its Search Engine ranking.

About half (46%) reported ‘refreshing page content’ as a regularly used method - although a useful ingredient, it seems slighted elevated when contrasted with the literature. Surprisingly, few cited ‘link development’ as an applied method to carry out SEO despite the exhortations of writers. While it has been suggested that techniques should be used in tandem in pursuit of optimal SEO, the variability in the logic and operation of each Search Engine’s algorithm means differing techniques may be needed to achieve effectiveness across each Search Engine.

4.5 Paid Search Strategy Techniques

When SMEs were asked about their most used Paid Search technique, Paid Placement was ranked first (see Table 1). Given the dominance of Google, this finding echoes the literature’s claim that paid placement is Google’s only Paid Search offering. The much smaller numbers for Paid Inclusion (9%) and Paid Submission (9%) appears to reflect the less widespread use of Search Engines like Yahoo and MSN who offer both types of Paid Search.
The significant number that does not use any Paid Search technique (29%) most likely reflects the earlier finding that reported SEO as the more used and effective technique. The reasonable level of Local Search is supportive of Laffey’s (2007) observation that this method assists the smaller business compete with the bigger players on a global scale and helps target specific geographic markets.

4.6 Measuring SEM

4.6.1 SEO Measurement

Traffic measurement was listed as the most used method for measuring SEO campaigns (see Table 2), concurring with the views of Zhang and Dimitroff (2005) and Friesen (2007). Traffic measurement is the most popular measure for SEO in practice, using tools such as Google Analytics. This was followed by ‘measuring link popularity’, as recommended by Schultz and Fristedt (2005). Disappointingly, ‘setting baselines and measuring the impact’ did not rate highly, despite its advocacy (Schultz and Fristedt, 2005). This technique should be more widely deployed since it is a true indication of a ranking on the SERP at one point in time, versus where it is at a future time having carried out SEO improvement measures. Worryingly, nearly a quarter said they do not use any method to assess their SEO initiatives. Given the current spend on SEM in Ireland and the expected growth of this form of online advertising (Taylor, 2007), this is a matter for concern.

4.6.2 Paid Search Measurement

On measuring Paid Search campaigns, firms indicated that both Traffic Measurement and CTR were jointly the most used methods (see Table 3). Fewer use Conversion Rate despite the view by Smith (2002) that conversion metrics are a very important tool for estimating ROI. Smith discusses how traffic measurement used to be the most appropriate measure, but the consumer conversion rate is now considered to be more important, where traffic measurement is more useful for analysing SEO initiatives (Zhang and Dimitroff, 2005; Friesen, 2007).

Reviewed at length in the literature are the large numbers of Websites that do not measure their Paid Search initiatives (Weischedel et al, 2005; Fain and Pedersen, 2006). This corresponds with the significant proportion here (37%) that indicates that they do not use any metrics to estimate their Paid Search efforts. This finding may be explained by the recognition throughout the literature that Paid Search lacks sound measurement tools (Weischedel et al, 2005). Nonetheless, implementing a Paid Search campaign and not measuring its outcome is futile.
4.7 Effectiveness of SEM Measurement

Opinions are divided evenly regarding how firm’s believe they are effectively measuring their SEM activity. Nearly half believe that they are effectively or very effectively measuring it, however the remainder are not at all happy with their current SEM metrics (see Figure 3). This view compounds the argument in the literature regarding the availability of effective tools and justifies why so many are discontented with their current arrangement (Fain and Pedersen, 2006).

Figure 3: Effectiveness of SEM Measurement.

As argued by Smith (2002), businesses using relevant metrics have a greater opportunity to achieve a competitive advantage. Metrics also provide valuable information about consumer’s online behaviour. Clearly much needs to be done to improve the satisfaction of firms in assessing the usefulness of their SEM effort.

4.8 Ethics and SEM

Respondents overwhelmingly agree that ethical considerations are a very important part of both their SEO and Paid Search campaigns. In concordance with this belief, there was general consensus that firms would have ethical issues in commissioning ‘all forms’ of Paid Search. Respondents strongly reported that they would only commission Paid Search that was transparent to users. These are very positive findings and would suggest that SMEs for the most part act ethically when undertaking their SEM and have given consideration to ethical issues.

Moxley et al (2004) view Paid Search advertising as misleading for information searchers as Search Engines often do not disclose the difference between paid and organic results. So while the most frequently subscribed to Search Engine firm, Google, (who clearly display the difference between sponsored results and organic results on the SERP) is used, searchers may have some confidence that firms are acting morally. Nevertheless, as Search Engine marketers change their Search Engine provider and develop their strategies (say to less transparent operators like MSN), an ethical policy should be carefully considered.

SMEs were more indifferent to whether their users were aware that they were using Paid Search. However, to a great extent it is outside of their control since the Search Engine firms design and deliver the SERPs. While Search Engine marketers might lobby those Search Engines that do not disclose the difference between paid and free Search Engine results, it is really up to consumer watchdogs and government agencies to ensure consumer rights are protected. Since there are a significant number of respondents who were unaware of what constitutes ethical SEO, these bodies have an important role in educating firms on the consequences of using third party marketing practices that are of questionable ethical merit and of other SEO practices such as Web spamming and unethical link development.

4.9 Challenges in SEM

When firms were asked about the main challenges encountered when undertaking SEM, the greatest one reported was competing for and achieving a high rank on the SERP. While earlier firms felt they were carrying out their SEO effectively, it still reflects the imperative for any SEM, which is to deliver users to their Website. A related challenge widely cited was the shortage of measurement tools available for SEM. Closely related to achieving a high ranking is the respondent’s concerns with the Search Engines algorithm. This will always be an ongoing challenge for marketers as Search Engines never disclose their ranking algorithms.

A significant number cited knowledge of SEM practices in general as an obstacle, indicative that much remains to be understood about this form of marketing for SMEs in Ireland to fully realize the opportunities it offers to their broader marketing initiatives. Dedicating time, budget and resources were listed by many respondents as a barrier to adopting SEM. This finding again illustrates SME’s lack of knowledge of SEM. Given the projected growth of SEM as a marketing medium - dedicating time, budget and resources is a crucial element in carrying out any SEM strategy. Managing keywords remain a major element for both Paid Search and SEO campaigns. In contrast, unethical competitor practices (click fraud, Web spamming and unethical
link development) are not a major concern for SMEs.

5 CONCLUSIONS

SEM is a growing strategy that firms use to improve the visibility of their Website so that searchers are delivered to the company Website to achieve their informational or commercial goals. The findings here suggest that Irish SMEs plan to spend more time and resources on SEM in the future. It is also clear that SEM strategies need to be more formal in nature, in line with recommendations made elsewhere. While it is to be expected that an important business initiative begins in a casual and experimental manner, very quickly the importance of aligning the effort with organisational strategy becomes evident. A fundamental mission of a SEM strategy, and clearly identified by respondents in this study, is to achieve a top position on the first page of the SERP.

The study also reveals that the great majority of respondents use SEO or a combination of SEO and Paid Search. Despite the general lack of comprehensive tools to measure SEM effectiveness, firms seem to have a clear preference for SEO at this time and perceive it to be more effective. To reinforce the perceptions of SMEs it is evident that SEM initiatives need to be measured with appropriate tools so that outcomes can be fed back into the strategic decision making process. What also emerges is that firms need to increasingly consider the behaviour and geographic location of searchers as an integral part of their SEM strategy.

An encouraging finding from the study is that Irish SMEs adopt a positive ethical posture when undertaking SEM. On a cautionary note however, since Paid Search is, for now, less widely employed, the more obvious ethical problems with it do not present themselves. Vigilance to give users transparency in recognizing paid versus free search results must be part of a firm’s SEM strategy if searchers are to remain confident in the integrity of Search Engine operatives.

REFERENCES


SHORT PAPERS
AN OPTIMIZATION METHOD FOR REDEMPTION AND DUE DATE MATCHING IN ASSIGNMENT OF ELECTRONIC RECEIVABLES BY USING INTEGER LINEAR PROGRAMMING

Toshiyuki Moritsu  
Systems Development Laboratory, Hitachi Ltd. 890, Kashimada, Saiwai-ku, Kawasaki, Kanagawa 212–8567, Japan  
toshiyuki.moritsu.xt@hitachi.com

Norihisa Komoda  
Graduate School of Information Science and Technologies, Osaka University 2–1 Yamadaoka, Suita, Osaka 565-0871, Japan  
komoda@ist.osaka-u.ac.jp

Keywords: Electronic Receivables, Integer Linear Programming, Finance, Due Date Matching, Optimization.

Abstract: This paper shows an optimization method for redemption and due date matching which assigns the receivables to the payments date under the pre-defined constraints which related companies specified. Our proposed method determines the pairs of the receivables and the payment with proper new settlement date which closes the fluid assets of companies to their target amounts by extending/shortening the redemption/due date. This paper shows that this matching program is categorized in integer linear programming. By applying this matching mechanism, transferors could utilize credit of issuers and also reduce fluid assets for payments. Effectiveness of our optimization method is shown by executing simulation which emulates the issuing and receiving status of receivables in Japanese companies.

1 INTRODUCTION

In recent years, various countries have made progress in the development of systems for handling credit transactions electronically, including EBPP/EIPP (Electronic Bill/Invoice Presentation and Payment) in the United States (Fairchild, 2003) and e-billing systems in South Korea (KFTC, 2002). Japan is now also expecting an increase in this type of business, and in June 2007 the Japanese parliament approved the Electronically Receivables Legislation (FSA, 2005) which is due to go into force at the end of 2008. Electronically receivables are a new form of credit whereby electronic registrations at organizations that register electronic receivables become only requirement for the generation and transfer of credit, and are expected to utilized as a means for the exchange of nominative claims and credits to payment on which contract documents and printed bonds are based (Ikeda, 2006).

The assignment of claims allocated to the payment of electronic receivables owned by a business has the merit of allowing the transferor to utilize the remitter’s credit rating (Oogaki, 2006). Specifically, when the transferor (company B) owns electronic receivables issued by a issuer (company A), even if company B has a poor credit rating and the electronic receivables issued by company B have not been received by the transferee (company C), then if we suppose that company A has a high credit rating then company B can pay by transferring credit issued by company A. This supplementation of reliability by a third party is particularly useful as a way of utilizing the credit rating of a parent company in the financing of subcontractor corporations in business affiliations that account for approximately 60% of such arrangements in Japan (JSBRI, 2007).

However, it has been pointed out that previous nominative claims and credits to payment are limited in terms of the opportunities for utilizing one’s credit in the assignment of claims. One possible reason for this is that the payment conditions (amount payable and due date) and credit redemption conditions are not exactly same. Since electronic receivables make it easy to rewrite the conditions electronically, it is relatively simple to accommodate differences in sums by splitting the electronic receivables. This has actually been implemented in the book entry transfer system for short term corporate bonds in Japan. On the other hand, in cases where the due dates do not match, it is necessary to adjust the gains and losses of the inter-
ested parties (issuer, transferor, transferee) before altering the conditions. This due date modification has the side benefits of allowing businesses to make effective use of surplus floating assets (referred to simply as “assets” in the following) and providing a means for supplementing shortfalls in assets. It is thus necessary to reach an agreement on items such as what the interest rate should be and by how much the term should be extended or contracted.

It is difficult to make an agreement between the interested parties regarding the change in due date by executing work flow between them because it can take much time to reach an agreement if consideration is given to factors such as changes of circumstances resulting from conflicting conditions.

Therefore in this study it is envisaged that making alterations to the due date of electronic receivables in an effective manner requires a mechanism whereby the participants first register assets fluctuation targets and variation conditions whereby changes in due date can be tolerated, and then combinations of payments and electronic receivables are determined to satisfy these constraints. For this study we therefore propose a matching scheme that optimizes combinations of payments and electronic receivables so as to achieve the best possible assets fluctuation targets for all participants. In particular, in this matching of electronic receivables and payments, the level of assets target achievement of the participants depends not only on which electronic receivables should be combined with which payments, but also on how the new due date for the redemption/payment of electronic receivables and payments is set and matched. We therefore formulate the issue of combining electronic receivables and payments by including the selection of a new due date whose selection range is determined by these combinations, and the results can be treated as a problem of integer linear programming. Also, by simulating the transaction of electronic receivables under conditions simulating the issue of corporate bills and accounts receivable in Japan, we first verify the extent to which the assignment of claim is promoted by the due date matching function. We also verify the optimization effects by comparing the optimized matching of electronic receivables and payments with some other combination methods.

2 METHOD FOR MATCHING THE DUE DATES OF ELECTRONIC RECEIVABLES

In this section we discuss a method for matching electronic receivables and payments based on the prior registration of due date alteration conditions which is a premise of this study. The processing flow is illustrated in Fig. 1. First, the participants register their respective asset fluctuation targets and due date adjustment criteria (1). The payment registrations are then accepted (2), the combination of electronic receivables and payments is determined based on these information (3), and approval is obtained for the matching results. The due date adjustment criteria are assumed to consist of a range of possible due dates for redemption payments, and interest conditions. In the following, we will discuss the asset fluctuation targets specified by the participants, the possible range of redemption/payment due dates, and the interest conditions.

- Asset Fluctuation Targets.
  Each participant specifies when and how much they would like the current assets to fluctuate. For example, the estimation of future changes in current assets is prepared from predictions of the payment and receipt of electronic receivables (and other settlements), and by setting the target holdings of current assets at each future timing, the fluctuation targets are set according to the difference between the target holdings and estimated transitions.

- Possible Range of Redemption/Payment due Dates.
  If necessary, the participants specify a range for the extension or contraction of payment due dates and/or redemption due dates.

- Interest Conditions.
  The participants specify a lower limit of interest to be accepted in cases where the redemption due date is extended or the payment period is reduced, and conversely an upper limit of interest to be paid in cases where the redemption due date is brought forward or the payment due date is extended. Note that the interest is set according to the debtor’s credit rating.
3 FORMULARIZATION OF THE PROBLEM OF MATCHING ELECTRONIC RECEIVABLES AND PAYMENTS

In this section we will devise a formula for the problem of matching electronic receivables and payments, and we will show that this problem can be expressed as a problem in integer linear programming (Schrijver, 1986)(Aardal et al., 2005).

The problem of matching electronic receivables and payments is expressed using the following notation:

**Constants.**

- \( R_{ijm} \): The m-th electronic receivable issued by issuer i and held by transferor (current holder) j.
- \( D(R_{ijm}) \): Redemption date of electronic receivable \( R_{ijm} \).
- \( DE(R_{ijm}) \): Upper limit of the extension of the redemption date of electronic receivables \( R_{ijm} \) set by issuer i.
- \( DS(R_{ijm}) \): Upper limit of the reduction of the redemption date of electronic receivables \( R_{ijm} \) set by issuer i.
- \( V(R_{ijm}) \): Redemption sum of electronic receivable \( R_{ijm} \).
- \( P_{jkn} \): Payment sum of electronic receivable \( R_{ijm} \) to be paid/received by transferor j.
- \( D(P_{jkn}) \): Payment date of payment \( P_{jkn} \).
- \( DE(P_{jkn}) \): Upper limit of the extension of the payment date of payment \( P_{jkn} \) set by transferor k.
- \( DS(P_{jkn}) \): Upper limit of the reduction of the payment date of payment \( P_{jkn} \) set by transferor k.
- \( V(P_{jkn}) \): Payment sum of electronic receivable \( R_{ijm} \) to be paid/received by transferor j.
- \( IP(i, R_{ijm}, f) / IR(i, R_{ijm}, f) \): Upper/lower limit of interest to be paid/received by issuer i in cases where the redemption date of electronic receivable \( R_{ijm} \) is extended/contracted to date \( f \). Issuer i decides the upper/lower limit of interest according floating assets needs.
- \( IP(j, R_{ijm}, f) / IR(j, R_{ijm}, f) \): Upper/lower limit of interest to be paid/received by transferor j in cases where the redemption date of electronic receivable \( R_{ijm} \) is extended/contracted to date \( f \). Transferor j decides the upper limit of interest according her/his floating assets needs and the lower limit according to credit rating of issuer of \( R_{ijm} \) and transferors (who transfer \( R_{ijm} \) before transferor j).

**Variables.**

- \( v(R_{ijm}, P_{jkn}, f) \): Transferred sum whereby electronic receivable \( R_{ijm} \) is allocated to payment \( P_{jkn} \) at a new due date \( f \).

The Objective function of this matching problem can be expressed as shown in Formula 1.

**Objective Function.**

\[
\text{Max}(\Sigma_{jk} v(R_{ijm}, P_{jkn}, f) \cdot (D(R_{ijm}) - D(P_{jkn})) - v(R_{ijm}, P_{jkn}, f))
\]  

(1)

The target of this problem is to make the best possible effort to ensure that the fluctuation targets of all participants are satisfied by the fluctuation of current assets resulting from changes to the due dates of electronic receivables and payments. To achieve this goal, it is preferable to match the combinations of payments and electronic receivables having longer discrepancies of due dates with the greatest transferred sums, within the range of the following constraints. As a result, the target function can be expressed as the sum total of the values obtained by multiplying the transferred sums by the discrepancies between the redemption due dates and payment due dates, and can be expressed by formula 1.

Meanwhile, the constraint formulae can be expressed by Formulae 2 through 15 as follows:

**Constraint formulae.**

\[
v(R_{ijm}, P_{jkn}, f) \geq 0
\]  

(2)

\[
V(R_{ijm}) \geq \Sigma_{jk} v(R_{ijm}, P_{jkn}, f)
\]  

(3)

\[
V(P_{jkn}) \geq \Sigma_{im} v(R_{ijm}, P_{jkn}, f)
\]  

(4)

\[
C_{xe} \geq \Sigma_{jk} v(R_{ijm}, P_{jkn}, f) + \Sigma_{j} v(R_{ijm}, P_{jkn}, f)
\]

where \( C_{xe} \geq 0 \) & \( D(R_{ijm}) \leq e < f \) & \( f < e \leq D(P_{jkn}) \)  

(5)

Figure 1: Process of Due Date Adjustment between Electronic Receivables and Payments.
\[
C_{xe} \leq \sum_{jkmn} v(R_{xjm}, P_{jkn}, f) - \sum_{jkmn} v(R_{jkm}, P_{jkn}, f)
\]

where \( C_{xe} < 0 \) & \( f < e \leq D(R_{xjm}) \) & & \( D(P_{jkn}) \leq e < f \)

\(
v(R_{jkm}, P_{jkn}, f) = 0
\)

where \( f \leq D(R_{jkm}) \) & & \( D(P_{jkn}) \leq f \) & & \( D(R_{xjm}) < D(P_{jkn}) \)

\( f < DS(R_{jkm}) \) & & \( f > DE(R_{jkm}) \) & & \( f < DS(P_{jkn}) \) & & \( f > DE(P_{jkn}) \)

Constraint formulae 7–15 define the possible range of combinations of electronic receivables and payments. Specifically, the range for which reverse combinations are not possible is defined as a transferred amount of 0 in formula 7, and the applicable ranges are specified by formulae 8–16.

Range 8 and range 9 specify that electronic receivables and payments cannot be matched unless the new due date after modification is between the redemption dates of the electronic receivables and the date on which the payment is due.

Range 10 specifies that electronic receivables and payments cannot be matched unless the new due date after modification is within the redemption dates of the electronic receivables and the possible range of payment due dates of the payment.

Constraint formula 5 shows that when the fluctuation target of the current assets of participant \( x \) on day \( e \) is increased, the increase target of participant \( x \) on day \( e \) is greater than the sum of the total amount of electronic receivables drawn by participant \( x \) for an extension astride day \( e \) and the total amount of payments in which participant \( x \) is the transferee for a reduction astride day \( e \) (a participant’s current assets are increased by the extension of redemption due dates on electronic receivables drawn by the same participant, or by the reduction in payment due date of payments received by the participant).

Constraint formula 6 is the converse of constraint formula 5, and shows that when the fluctuation target of the current assets of participant \( x \) on day \( e \) is decreased, the reduction target of participant \( x \) on day \( e \) is less than the sum of the total amount of electronic receivables drawn by participant \( x \) for a reduction astride day \( e \) multiplied by the negative value of the total amount of payments in which participant \( x \) is the transferee for an extension astride day \( e \) (a participant’s current assets are reduced by the reduction of redemption due dates on electronic receivables drawn by the same participant, or by the extension in payment due date of payments received by the participant).

The constraint formulae consist of the following three types:

1. Constraints in which the possible range of transferred sums is predetermined (formulae 2–4)
2. Constraints in which fluctuations of current assets arising from transfers associated with changes of due dates are kept within the fluctuation target (formulae 5–6)
3. Constraints in which the electronic receivables and payments that can be combined are predetermined (formulae 7–16)

Constraint formula 2 imposes the condition that the transferred sum must be positive.

Constraint formula 3 imposes the condition that the sum total of sums transferred when electronic receivables are transferred in separate parts does not exceed the sum of the original electronic receivables (but not necessarily equal since the splitting of electronic receivables may leave parts that are not suitable for transfer).

Constraint formula 4 imposes the condition that the sum total of payments made by the transfer of electronic receivables does not exceed the total amount of be paid (but not necessarily equal since new issues and combinations are made in cases where the payable sum cannot be fully allocated by the transfer of electronic receivables alone).

Constraint formula 5 shows that when the fluctuation target of the current assets of participant \( x \) on day \( e \) is increased, the increase target of participant \( x \) on day \( e \) is greater than the sum of the total amount of electronic receivables drawn by participant \( x \) for an extension astride day \( e \) and the total amount of payments in which participant \( x \) is the transferee for a reduction astride day \( e \) (a participant’s current assets are increased by the extension of redemption due dates on electronic receivables drawn by the same participant, or by the reduction in payment due date of payments received by the participant).
due to this reduction.

Range 12 relates to the case where the redemption due date comes after the date on which the payment is due, and is the converse of range 11 where the payer/payee relationships of interest payments by the debtor, transferor and transferee are reversed.

Range 13 specifies that when there is an increase in the fluctuation target for current assets on day \( e \), it is not possible to perform matching with electronic receivables issued by the issuer involving a change in the redemption due date so as to reduce the redemption due date astride day \( e \) (a change that depletes current assets). Since this means it is possible to exceed the target if only one of the combinations is agreed upon, in this formulation the fact that fluctuation targets cannot be exceeded is added as a premise to this constraint.

Range 14 relates to the converse of range 13 in cases where the current formula fluctuation target of the issuer is reduced.

Ranges 15 and 16 are the same as ranges 13 and 14 where the issuer imposes limits on the redemption due data by means of restrictions on changes to the payment due date in transferee \( k \).

With regard to the definition contents of the above target functions and constant formulae, first, target function 1 is the maximization of a primary function with \( v(R_{jm}, P_{kn}, f) \) as a variable, and it takes an integer value. Also, the constraint formulae are all first-order inequalities that take \( v(R_{jm}, P_{kn}, f) \) as a variable. This problem can therefore be classified as an integer linear programming problem.

4 EVALUATION

The effects of changes in due date and the effects of optimizing the combinations of electronic receivables and payments were verified by simulation. This section discusses the preconditions under which the simulation was conducted, and then presents the measurement results.

4.1 Preconditions

There is currently no statistical information relating to the issue of electronic receivables. We therefore performed the simulation by assuming conditions for the issue of electronic receivables based on financial information from Japanese businesses. Table 1 shows the financial information and the conditions for the issue of electronic receivables assumed in this simulation. The financial information was sourced from corporate statistics published by the National Tax Agency and from settlement trends for 2003 published by the Bank of Japan, including the average sales figures for Japanese corporations, payable liabilities (accounts payable, bills payable), received credit (accounts receivable, bills receivable), average sum of bills cleared, and average sum of accounts receivable. The conditions for the issue of electronic receivables were assumed based on this financial information. Specifically, we made assumptions regarding the average redemption period of electronic receivables, the average frequency of issue and the average sum. The respective calculation formulae are shown below.

- Average redemption period of electronic receivables = \( \frac{\text{receivable credit}}{\text{sales}} \times 365 \)
- Average number of electronic receivables issued = \( \frac{\text{bills receivable}}{\text{average sum of bills cleared + accounts receivable}} \times \frac{\text{average sum of accounts receivable}}{365} \)
- Average sum of credit = \( \frac{\text{receivable credit}}{\text{bills receivable}} \times \frac{\text{average sum of bills cleared + accounts receivable}}{\text{average sum of accounts receivable}} \)

In the simulation, the number of companies was taken to be 260 (one thousandth of the actual number of businesses), and measurements were performed by repeating the transactions over two years. In real situations, not necessarily all the credit is replaced with electronic receivables, and not necessarily all the electronic receivables are subject to being transferred, so the simulation was performed by making a few changes to the ratio of transferable sums with regard to the credit sums of the electronic receivables belonging to a business. The simulation environment parameters were as follows: MPU: Xeon\(^1\) 2.8 GHz, Memory: 3 GByte, Windows XP\(^2\), JDK 1.6.0 01_j13\(^3\), LpSolve 5.5.0.10(Berkelaar et al., 2004). In the evaluation results shown in the next section, measurements were also performed by varying some conditions of the other parameters (variation in redemption periods of electronic receivables, variation in frequency of issue of electronic receivables, variation in monetary value of electronic receivables, number of companies simulated), but the effects of these changes were smaller than those of the parameters shown in Table 1 and thus these results are omitted.

\(^1\) Xeon is a registered trademark of Intel Corporation.
\(^2\) Windows XP is a registered trademark of Microsoft Corporation.
\(^3\) Java is a trademark of Sun Microsystems, Inc.
Table 1: Financial Statement of Average Japanese Company and Assumed Issue Condition of Electronic Receivables.

<table>
<thead>
<tr>
<th>Description</th>
<th>Amount (k¥)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Total Sales</td>
<td>20,483</td>
</tr>
<tr>
<td>Total Receivables</td>
<td>82,887</td>
</tr>
<tr>
<td>Account Receivables</td>
<td>66,287</td>
</tr>
<tr>
<td>Note Receivables</td>
<td>16,600</td>
</tr>
<tr>
<td>Average Amount of Account Receivables</td>
<td>2,000</td>
</tr>
<tr>
<td>Average Amount of Note Receivables</td>
<td>4,000</td>
</tr>
<tr>
<td>Assumed Average Redemption Period of Electronic Receivables (Day)</td>
<td>59</td>
</tr>
<tr>
<td>Assumed Average Issue Cycle of Electronic Receivables (Times/Day)</td>
<td>0.102</td>
</tr>
<tr>
<td>Assumed Average Amount of Electronic Receivables (k¥)</td>
<td>2,222</td>
</tr>
</tbody>
</table>

Figure 2: Ratio of Payments by Transfer.

4.2 Measurement Results

We will first use the due date modification function to investigate the extent to which payments are promoted by transfers. Next, by optimizing the combinations of payments and electronic receivables, we will verify the extent to which it is possible to achieve the participants’ fluctuation targets. Finally, we will investigate the computational load required for optimization.

The graph in Fig. 2 shows the ratio of all payments in which payment was made by assignment of claim. The horizontal axis shows the number of days by which the payment due date and redemption due date can be varied in either direction. The multiple measurement results in this graph correspond to measurements made while varying the criteria regarding the extent to which the electronic receivables owned by a business are transferred on a monetary basis. This evaluation shows the results obtained when the matching of electronic receivables and payments is optimized.

In these results, the case where the modifiable date is 0 corresponds to the case where no changes can be made to the redemption due date of the electronic receivables. In this case, the ratio of the payment that can be assigned in the transfer of electronic receivables is about 3.3% of the total. On the other hand, when the redemption due date and payment due date are set with the ability to be varied forwards or backwards by up to 3 days, the ratio of the payment in the transfer increases to 16.7% (about 5 times larger) when all the electronic receivables are transferred. Also, even when the ratio of the electronic receivables to be transferred is reduced to about 1/32 of the total electronic receivables, an increase of 12.3% (about 3.7 times) is seen. These results show that by providing a framework for making adjustments of a few days in the redemption due date or payment due date, each business can make a relatively large improvement to the ratio of payments made by transfers.

The optimization results are shown next. Figure 3 shows the extent to which it is possible to achieve the fluctuation targets of current assets set by the participants for each method of matching a number of electronic receivables and payments. The specific matching methods used to make these measurements were as follows:

- Randomly ordered Matching.
  Matching is performed in random order from among candidates consisting of electronic receivables and payments that are capable of being combined.

- Matching in Order of Effects on the Level of Achievement.
  Matching is performed in order of the magnitude of how the achievement of the fluctuation targets is affected by candidates consisting of electronic receivables and payments that are capable of being combined.

- Matching based on Optimal Pattern Searching.
  Matching is performed by searching for optimal patterns that are closer overall to the fluctuation targets.

Figure 3: Ratio of Adjusted Asset to Target.
The horizontal axis in the graph of Fig. 3 shows the number of days by which the payment due date and redemption due date can be varied in either direction. The vertical axis shows the extent to which the fluctuation targets preset by each business could be achieved. With regard to these measurements, the fluctuation targets of the businesses were set randomly in the range of the sums of electronic receivables owned by these businesses. Figure 4 compares the improvement rate of optimization between random matching and matching based on the effects on the level of achievement. These results show that although there is some degree of variation when there are few modifiable days, the average ratio of improvement is approximately 18.1% for matching based on the effects on the level of achievement, and approximately 19.6% for random matching.

Next, the computation time needed to optimize the combinations of electronic receivables and payments is investigated using the graph shown in Fig. 5. In this graph, the number of businesses is shown on the horizontal axis and the time required for the optimization computations per day is shown on the vertical axis. In a simulation of 260 companies, the computation time required for optimization was 113 seconds per day on average. This remains future work to make this method in practice because this simulation done with 1/1000 of actual number of companies and the computation time grows exponentially according to increase the number of companies. We discuss this problem in section 5.

Finally, Table 2 shows the extent to which the electronic receivables are fragmented by the matching of due dates. The number of fragments is a value that shows on average how many electronic receivables the original electronic receivables are divided into at the time of redemption. The average number of fragments was 2.08 when matching was performed in random order, 1.38 when matching was performed in order of the effects on the level of target achievement, and 2.06 when matching was performed based on optimal pattern searching. The reason for the small number of fragments obtained when matching in order of the effects on the effects on the level of target achievement is thought to be because in this algorithm, matching is performed preferentially on groups of electronic receivables and payments involving larger sums and greater differences in due dates, and these larger sums suppress the fine fragmentation of electronic receivables.

5 CONCLUSIONS

In this study, we have proposed a scheme for optimizing the matching of redemption due dates and payment due dates for electronic receivables to promote the transfer of electronic receivables. Specifically, we have demonstrated an optimization scheme that makes every effort to make the fluctuations in the current assets of each business (caused by changes of due date) approach their fluctuation targets based on criteria specified by each business regarding the fluctuation targets of current assets, the range of alterations to due dates, and the rates of interest. We have also shown that this problem can be classified as a type of integer linear programming problem. By performing simulations based on the financial circumstances of average Japanese businesses, we have shown that this technique is capable of promoting transfers involving changes of due date and improving the degree to which fluctuation targets are achieved by optimization.

One issue for further study is the problem of electronic receivables being finely fragmented by the optimization process. In the method of this study, if the fluctuation targets of the participants’ assets are im-
proved, then the electronic receivables can be arbitrarily fragmented and a number of payments will be assigned to a number of new due dates. This fragmentation of electronic receivables leads to increased administration costs, so to actually put this method into practice, it is important to investigate how to control the fragmentation of electronic receivables.

A second issue is that of partitioning the optimization regions. In this study, optimization was performed by using a single target function to represent the criteria of all the companies concerned, but this is inefficient with regard to increasing the scale of businesses to which the method is applied. In practice, there is considered to be some degree of locality in the transaction relationships between businesses, so if groups of businesses can be split into suitable ranges, then it should be possible to split the optimization problem into multiple sub-problems with fewer variables and constraints. In this way it should be possible to reduce the computational cost and speed up the computation time by employing parallel processing.

A third issue is that not necessarily all of the optimal matching results demonstrated by this algorithm are the best matching results from the viewpoint of each individual business. For actual operations, an important issue is therefore to somehow present alternative proposals in cases where a business rejects the matching results.

REFERENCES


E-LEARNING FOR NEW GRADUATE EMPLOYEES
Another Function of e-Learning for New Graduate Employees of Japanese and Korean Companies

Jiro Usugami
Department of Economics, Oita University, 700 Dannoharu, Oita City, 870-1192, Japan
usugamij@cc.oita-u.ac.jp

Keywords: New graduate employee, human capital development, employee retention, pre-entry training program.

Abstract: Japanese companies tend to consider that e-learning is important for a reason other than human capital development in new graduate employee training and development programs, namely for improving employee retention. Two remarkable features of the Japanese employment system are the seasonally scheduled hiring of new graduates and company commitment to the T&D programs for new graduate employees before and after their official hiring. Today, large- and medium-sized Japanese companies make their offers to candidates nearly one year prior to their graduation from university or college. Many of these companies also provide various T&D programs, including e-learning, for the potential new graduate employees until their graduation. We call these “pre-entry training” programs. They are distinguished from internship programs in that pre-entry training programs are provided only for those students who are due to be officially hired the next fiscal year. Besides Japanese companies, a number of foreign companies in Japan and some Korean companies provide pre-entry training programs that utilize e-learning. This study examines, based on case studies, the hypothesis that e-learning in the pre-entry training programs for new graduate employees has the function of retention as well as human capital development.

1 INTRODUCTION

Many Japanese companies utilize e-learning in the training and development programs for new graduate employees, before and after their official hiring.

New graduate hiring is done seasonally in Japan. Large- and medium-sized companies make their offers to candidates nearly one year prior to their graduation from university or college. During this long period until graduation, many of these companies provide various T&D programs, including e-learning, for the potential new graduate employees, which we call “pre-entry training” programs. They are distinguished from internship programs in that pre-entry training programs are provided only for those students who are due to be officially hired the next fiscal year.

Besides Japanese companies, a number of foreign companies in Japan and some Korean companies also provide pre-entry training programs that utilize e-learning.

This study examines the hypothesis that e-learning in the pre-entry training programs has the function of retaining potential new graduate employees as well as developing human capital, based on case studies.

In Japan and Korea, there is intense competition among companies hiring new graduates, which makes employee retention a critical issue. Japanese companies need to retain potential new graduate employees for about one year prior to official hiring, and tend to consider that e-learning in the pre-entry training programs is an effective retention strategy.

This study comprises the following chapters. Chapter 2 provides an overview of the new graduate hiring schedule in Japan and typical e-learning menus in the pre-entry training programs. Chapter 3 presents a theoretical model and hypothesis on the function of e-learning in the pre-entry training programs. Chapter 4 analyzes the objectives of companies utilizing e-learning in the pre-entry training programs for potential new graduate employees, based on case studies. Chapter 5 provides conclusions.
2 HIRING SCHEDULE & E-LEARNING

Large- and medium-sized Japanese companies tend to hire a certain number of new graduates every April when the new fiscal year begins. By the time they are officially hired, many of these new graduate employees have experienced a variety of e-learning courses provided by the companies.

This chapter presents an overview of the new graduate hiring schedule in Japan and typical e-learning menus in the pre-entry training programs.

2.1 New Graduate Hiring Schedule in Japan

The new graduate hiring season starts more than one year prior to graduation. Students graduate in March and the Japanese fiscal and academic year begins in April. Hiring tends to be scheduled as follows.

First, from October to December, many Japanese companies announce their new graduate recruiting to students in their third year of university or college. From December to March of the following year, job fairs are held in several big cities in Japan. Companies also provide career guidance and conduct on-campus recruiting. From January to April, written tests and job interviews are conducted for candidates in their last year of school.

From April to June, companies select the new graduate employees for the next fiscal year, and make their offers to the candidates. During a period of about six months from October of the students’ last year to March when they graduate, many of these companies provide pre-entry training programs that include e-learning courses, events, classroom instruction, camp training, and on-the-job training for those students who are due to be officially hired after graduation. According to a survey conducted by the Institute of Labor Administration, over 60% of large- and medium-sized companies provided pre-entry training programs in 2005.

Then, on the first business day in April of the new fiscal year, initiation ceremonies are held and new graduates are officially hired.

2.2 Typical e-Learning Menus

E-learning in the pre-entry training programs has a variety of menus. Large enterprises tend to develop their own e-learning courses. Typical menus of outsourced e-learning are language education (English, Chinese, etc.), PC skills development (Word, Excel, PowerPoint, etc.), business skills development (bookkeeping, marketing, etc.), and business etiquette and communication (business greeting protocol, telephone etiquette, etc.). There are also orientation programs that include company overviews and support programs for acquiring business certificates.

3 MODEL AND HYPOTHESIS

What is the function of e-learning in the pre-entry training programs for new graduate employees?

Many large- and medium-sized Japanese companies provide Off-JT programs such as class instruction, role-play training and camp training after officially hiring new graduates. On the other hand, in the pre-entry training programs, e-learning tends to be easier to provide since the trainees are still attending university or college at various locations including overseas. Therefore, companies recognize the effectiveness of e-learning.

This chapter presents a theoretical model and hypothesis on the function of e-learning in the pre-entry training programs.

3.1 Theoretical Model

The principal objective of e-learning is human capital development. According to Becker (1975), the major idea underlying human capital theory is that an individual’s performance is determined by his/her stock of knowledge and productive skills.

Applying this theory to the case of new graduate employees, those who have accumulated some stock of knowledge and skills by the time they are officially hired would demonstrate good work performance from the outset. In addition, providing Off-JT programs after their hiring would be efficient and performance on the job would be high.

Figure 1 shows a basic human capital development model. Dotted and solid lines indicate, respectively, the performance of new graduate employees who experienced Off-JT after being hired, and those who experienced only OJT.

Figure 2 is a human capital development model for a case in which pre-entry training programs are provided. The broken line indicates the performance of new graduate employees who experienced pre-entry training before being hired and OJT afterwards. The movement of the dotted line indicates that pre-entry training shortens the time for Off-JT programs after hiring and improves its efficiency.

This model suggests that companies could accelerate the human capital development of new
graduate employees, improve the efficiency of T&D programs for them, and gain high performance sooner by making e-learning programs available to all trainees wherever they are located. However, the model does not refer to the relationship between e-learning and employee retention.

![Figure 1: Basic Human Capital Development.](image)

**3.2 Hypothesis**

With today’s expanded menus and contents for new graduate employees, e-learning has become an easy tool for developing the human capital of students who are due to start working after graduation. However, we believe that there is another reason why Japanese and Korean companies are utilizing e-learning in the pre-entry training programs, besides human capital development and improving the efficiency of T&D programs (Figure 3). We provide a hypothesis for our analyses.

![Figure 2: Pre-Entry T. & Human Capital Development.](image)

Hypothesis: Companies recognize that e-learning in the pre-entry training programs is important for retaining potential new graduate employees as well as accelerating human capital development and improving the efficiency of T&D programs for them.

**4 ANALYSES**

What are the actual objectives of the companies that utilize e-learning in the pre-entry training programs?

Usugami (2006) conducted a survey in 2005 and found that Japanese and Korean companies of 137 samples aimed to maintain the motivation of potential new graduate employees, improve their interpersonal skills and retain them by providing pre-entry training programs that include e-learning.

Hanada & GaiaX (2008) conducted a survey for companies that provided pre-entry training programs including SNS (Social Networking Service) in 2008. They reported that of those that replied, 24 companies recognized the prime importance of retaining potential new graduate employees, easing their anxiety, and building their interpersonal skills by providing activities for them until their graduation. Furthermore, Hanada & GaiaX revealed that these companies adopted SNS in the pre-entry training programs in order to build the interpersonal skills of potential new graduate employees, retain them, and maintain their motivation.

This chapter provides three case studies and discusses the company’s recognition of the function of e-learning in the pre-entry training programs.

**4.1 Case Study 1**

Case 1 is a large Japanese insurance company. They adopted e-learning in their six-month pre-entry training programs in October 2006. In their e-learning program, potential new graduate employees are given individual pages and provided with compulsory and optional menus. Compulsory menus are comprised of certificate acquisition courses such as security trading and financial planning, PC skills development courses such as Excel and PowerPoint, TOEIC courses, management overview, and business etiquette. Trainees who tend to lag behind are encouraged by emails from Human Resource Development staff. As an optional menu, video courses are provided for about 300 subjects. The most popular subjects among trainees are business communication, team building, time management, and marketing.

The company in Case 1 indicated that the objectives of adopting e-learning were to retain potential new graduate employees, average out their
knowledge and skills, develop human capital, and improve the efficiency of T&D programs after officially hiring them.

4.2 Case Study 2

Case 2 is a large Japanese service company that regards employee retention as an important issue. In July 2007, they adopted SNS for potential new graduate employees.

The company is building a specific community in the SNS, and provides CEO messages, overviews of management and group companies, and a variety of information. In addition, recruiters and young employees join the community as advisers.

This company adopted SNS in order to retain potential new graduate employees. They are asked to post their opinions about events held for them on the community board. The company utilizes their opinions for the next new graduate recruiting and pre-entry training programs.

4.3 Case Study 3

Case 3 is a large Korean manufacturing company. As in Japan, there is seasonally scheduled hiring of new graduates in Korea also. Large- and medium-sized companies tend to make their offers to candidates three to six months prior to their graduation from university or college. The period until official hiring is shorter than that in Japan, but the competition between Korean companies is just as intense, and potential new graduate employee retention is a critical issue.

According to an interview survey that we conducted in February 2008, this company provides e-learning courses, events, and mentoring in the pre-entry training programs for potential new graduate employees. Their e-learning menus comprise orientation and PC skills development in programming. The contents of orientation include overviews of management philosophy, company history, business model, and products and services.

The company in Case 3 mentioned that they recognized e-learning as an effective means of retaining potential new graduate employees, developing their basic business skills, and improving the efficiency of T&D programs for them (Table 1).

4.4 Discussion

E-learning is utilized as a tool to develop the human capital of employees and improve the efficiency of T&D programs for them.

In fact, the companies of the three case studies recognize that e-learning is important for developing the human capital of potential new graduate employees and improving the efficiency of T&D programs for them. Besides these known functions, the case studies show that Japanese and Korean companies expect e-learning to help improve employee retention when hiring new graduates. This feature was demonstrated by the survey of Hanada & GaiaX.

Japanese and Korean companies make their offers more than a few months prior to students’ graduation amid intense competition to hire new graduates, and provide pre-entry training programs until officially hiring them. As a result, companies tend to focus on e-learning’s function of retaining potential new graduate employees, besides the conventional functions of maintaining their motivation, developing their basic business skills, and improving the efficiency of T&D programs for them.

5 CONCLUSIONS

This study examined the hypothesis that e-learning in the T&D programs for potential new graduate employees has the function of retention as well as
human capital development, based on the recent survey and case studies. Many Japanese companies and some Korean companies provide T&D programs that include e-learning for potential new graduate employees prior to official hiring. These are known as pre-entry training programs. Today, e-learning has numerous menus and contents and is an easy means of providing pre-entry training programs to potential new graduate employees who are located in various places including overseas.

E-learning has been utilized for human capital development. We investigated another reason why companies adopt e-learning in their pre-entry training programs, based on the recent survey and three case studies. We found that companies tend to consider that e-learning in the pre-entry training programs is important for retaining potential new graduate employees as well as accelerating human capital development and improving the efficiency of T&D programs for them.

REFERENCES


Keywords: e-Supply chain, systems engineering, e-Business adoption, network organisation, structural equation modelling, SMEs.

Abstract: This paper reviews the available literature to identify theoretical and empirical gaps regarding e-Supply chain adoption among small and medium sized enterprises (SMEs). We argue that a generic e-Supply chain framework should employ appropriate operational and behavioural perspectives. We propose that e-Supply chain operation can succeed only when integrated with an efficient supply chain network and a strategic plan committed to e-Business. Survey data from 208 Malaysian SMEs are collected and Structural Equation Modelling (SEM) is employed to apply the proposed framework. More specifically, a set of three empirical models are examined to evaluate the validity and impact of supply chain strategy, e-Business adoption, and the interaction of these constructs on business performance. Our review suggests that much of the existing e-Supply chain adoption and implementation literature is not firmly grounded in theory. We have suggested that the systems engineering tradition of focusing on the interactions of technology, organizational structure, and personnel provides a useful framework for understanding the business performance of e-Business. The technology-organisational and people (TOP) dimensions are based on sound systems engineering principles that are widely recognised and accepted for improving manufacturing organisation. We suggest that these principles are equally valid for the e-Business oriented and virtual organisations. These expectations are supported by our empirical results. We find that measures based on the TOP conceptual scheme provide reliable and valid scales that are equally applicable to both e-Business and non-e-Business firms.

1 INTRODUCTION

Recent years have witnessed the worldwide adoption of e-Business for achieving cost savings, improving customer service, and promoting innovation, and taking advantage of new business opportunities (Wagner et al., 2003). Despite the burst of the dot-com bubble, companies are continuing to adopt e-Business operations. Martinsons and Martinsons (2002) suggest that the fear of lagging behind in adopting the Internet technology has rushed many firms to blindly engage in e-Business initiatives. As a consequence, many have done so without deriving much benefit. Despite huge investments in e-Business initiatives, academics and practitioners are questioning the value proposition of e-Business investments (Zhu et al., 2003).

Studies indicate that technology diffusion among knowledge intensive SMEs have been difficult (see for example, McCole and Ramsey; 2005; Ramsey et al., 2005). Chapman et al. (2000) argue that SMEs are lagging behind their larger counterparts in the use of the Internet. Other studies have found that SMEs are only half as likely to be using e-mail; for micro companies the figure is even smaller (Chapman et al., 2000). Poon and Swatman (1997) suggest that cost is the largest barrier for SMEs restricting the adoption of new technology. A recent report by Spectrum (2001) also supports the view that SMEs need to catch up with their larger
counterparts in adopting ICTs in the automotive components sector. The main issues identified include the excessive cost of e-technology and skill deficiencies in e-Business implementation.

This paper provides a critical review of the available e-Business literature to identify theoretical and empirical gaps. Based on our literature review, we identify operational and behavioural perspectives that form the basis of a theoretical framework for understanding e-Supply chain adoption and success. We show that these perspectives relate to the well-established “systems engineering” principles of technology, organisation, and people.

2 E-SUPPLY CHAIN: LITERATURE REVIEW

Operations management academics have always highlighted the strategic importance of operations, and its role in corporate success. The consideration of operation strategy is relatively as important in e-Business operations as in operating in traditional environments. However, evidence from the literature suggests that many companies have adopted e-Business without thinking through their strategic, operational and behavioural impacts (Marshall and Mackay, 2002; Gunasekaran et al., 2002; Dutta and Biren, 2001), which subsequently led to e-Business failure. This section considers the impact the Internet has on operational, and behavioural management perspectives and whether new strategic thinking is required in response to the powerful external forces that are re-shaping industry. This section also aims to sustain the significance of these perspectives by providing supporting evidence from the existing e-Business literature.

2.1 Operational Perspective

e-Business is important for the supply chain literature because of the increasing need to integrate activities and information flows and to optimise the processes not only at the single company level, but also at the level of inter-company processes (Landford, 2004; Lattimore, 2001; Stevens, 1989). The importance and role of web-based technologies to support company operations (e-Supply chain) is widely acknowledged by both practitioners and academics (Sanders and Premus, 2005; Porter, 2001; Skjoett-Larsen, 2000).

There has been extensive research investigating the impact of organisational factors on innovation and technology adoption (Fjermestad, 2003; Grandon and Pearson, 2004). The factors influencing Internet technology adoption within supply chain strategy can be classified in several ways such as internal and external environments, firm and individual conditions, and domestic and international involvement (Moini and Tesar, 2005; Lewis and Cockrill, 2002). The perceptions of management toward IT adoption are examined in many studies (Taylor and Murphy, 2004).

2.2 Behavioral Perspective

Technological sophistication of an organisation is considered an important factor for businesses’ e-Business adoption and implementation. There has been extensive research outlining important determinants of organisational factors on e-Business adoption (Tornatzky and Fleischer, 1990). The majority of organisational factors addressed involve such organisational characteristics as size, industry type and business scope (Zhu et al., 2004, 2006). However, there is a lack of study addressing the relationship between information orientation / asymmetry and technological innovation / integration on e-Business adoption (Hsieh et al., 2006).

From the behavioural perspective, Damodaran and Olpher (2000) have identified knowledge transfer, knowledge integration, and practical application of knowledge as the main elements for developing “external” capabilities. According to a study conducted by Caloghirou et al. (2004), the readiness, and openness towards knowledge sharing among business partnerships are important factors in improving business performance and encouraging the adoption of e-Business. Establishing knowledge management mechanisms and advantage knowledge assets is essential for successful technological and organisational innovation (Bong et al., 2004).

2.3 Performance Measurement

Marshall et al. (1999) define performance measurement as “… the development of indicators and collection of data to describe, report on and analyse performance”. Neely et al. (1995) see performance measurement as “the process of quantifying the efficiency and effectiveness of action”. Sanders and Premus (2005) argue that performance measurement is a complex issue that incorporates economics, management, and
accounting disciplines. Zhu et al. (2004) have stressed that an appropriate measurement system is essential to support a wide range of performance measures. Using Kaplan and Norton’s (2004) balance score card concepts, we have identified tangible and intangible performance measures to evaluate performance improvements (Hafeez et al. 2007). Based on the relevant literature (Eikebrokk and Olsen, 2005) we identify three domains of measures to examine the perceived benefits of e-Business adoption: Financial, Operational efficiency and Coordination. Within each domain, it is useful to categorise specific indicators under “operational”, and “behavioural” perspectives.

2.4 System Engineering Concept

Systems engineering may be defined as the science of analysing the behaviour of a system (or organisation) by studying the technology, policies and management procedures (or organizational structure) and the behaviour and attitudes of the people who make up of the organisation (Forrester, 1961; Parnaby, 1981; Towill, 1993). Many past and current management initiatives such as Total Quality Management (TQM) (Hafeez et al. 2006, supply chain management (Hafeez, et al. 1996), business process re-engineering (BPR) (Hammer and Champy, 1994) are based on systems engineering principles. Systems engineering distinguishes technology (T) and/or organisation (O) and/or people (P) dimensions (or TOP dimensions in short). Systems engineering emphasizes the interconnectedness of these dimensions, and suggests that change in one is very likely to have implications requiring changes in others.

e-Business operation might best be understood from the perspective of supply chain management. We would particularly draw attention towards Stevens’ (1989) supply chain management integration framework based on systems engineering principles. Stevens’ (1989) model also provides a consistent empirical support, which provides a good base for comparisons. Stevens (1989) has differentiated contributory factors for supply chain integration into the ‘hard’ issues (such as technology) and the ‘soft’ (e.g. relations, attitudes, etc). Numerous studies suggest that many companies have not yet fully realised the technological integration of the available office technologies and software tools such as Material Resource Planning (MRP), Distribution Resource Planning (DRP), and Enterprise Resource Planning (ERP). Stevens, as early as 1989, advocated that in order to achieve full integration (from a baseline to external; companies needed to focus on people dimensions internally as well as externally. This study argued the applicability of Stevens’ (1989) integration framework in today’s business environment where companies want to move from a traditional business to e-Business. Therefore, the identified dimensions, namely technology, organisation, and people (TOP) are well suited for studying the success of e-Supply chain adoption.

Table 1: Incorporation of technology, organisation and people dimensions within each identified factor.

<table>
<thead>
<tr>
<th>Variables examined</th>
<th>Investments for supply chain system</th>
<th>Integration of operating and planning database</th>
<th>Standardised and customised information</th>
<th>Information sharing and distribution</th>
<th>Organisational structure</th>
<th>Standardised supply chain practices and operations</th>
<th>Integration of individual operations channel</th>
<th>Time based logistics solutions</th>
<th>Roles and responsibilities</th>
<th>Developing and maintaining relationships</th>
<th>Risk and rewards</th>
</tr>
</thead>
<tbody>
<tr>
<td>“Operational Perspective” Supply Chain Strategy</td>
<td>Technology innovation and integration</td>
<td>Information orientation and asymmetry</td>
<td>Adoptability of technology infrastructures</td>
<td></td>
<td>Organisational learning factors</td>
<td>Organisational support and value</td>
<td>Organisational knowledge management</td>
<td>Internal and external collaboration</td>
<td>Performance measurement</td>
<td>Readiness mindset of adoption</td>
<td></td>
</tr>
<tr>
<td>“Behavioural Perspective” E-Business Adoption</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Following the critique from the literature and gaps identified, it can be seen that the context of operational and behavioural management are still fit to investigate the success factor of e-Business adoption. Through a careful content analysis, elements have been identified which in the present author’s view contribute to e-Business research. They can be generally categorised under the well-
established operations research dimensions of technology, organisation and people (see Table 1).

3 THEORETICAL FRAMEWORK

We have argued that a successful e-Supply chain company needs to take into account “operational” and “behavioural” issues. The overarching theoretical framework is summarized in Figure 1, and includes supply chain strategy and e-Business adoption constructs.

![Figure 1: A conceptual framework for e-Supply chain adoption.](image)

Figure 1 illustrates that within each of these constructs are embedded the three systems engineering principles: “technology”, “organisation,” and “people” (or TOP). The framework illustrates that these constructs are interrelated, and therefore any change in one factor will have ramifications for others. We hypothesize that developments in each dimension (TOP) of each domain (supply chain strategy and e-Business adoption) are necessary for satisfactory business performance (BP).

Hypothesis H1: Supply chain strategy (SCS) is a significant determinant of business performance (BP)

Hypothesis H2 E-Business adoption (EBA) is a significant determinant of business performance (BP)

Hypothesis H3 Business performance (BP) is directly related to the level of mutual dependency (and alignment) between supply chain strategies (SCS) and e-Business adoption (EBA)

In the sections that follow, we will illustrate the utility of our approach by applying it to understanding variation in the business success of Malaysian SMEs.

3.1 Sample Selection

We also describe the results obtained of an empirical study applying these concepts to data collected from 208 Malaysian SMEs. Using confirmatory factor analysis techniques (i.e. structural equation modelling or SEM), we show the effects of operational, and behavioural adjustments to e-Business on business success. Malaysia has been developing its information highway capacity since late 1990s. This is realised by the investment of RM 40 Billion (approximately £ 5.9 billion) to establish Multimedia Development Corporation (MDC). The Multi-media Super Corridor (MSC) is one of the key initiatives of MDC (Low et al., 2000).

We have selected six industrial sectors that had previously been identified as the leading sectors in e-Business adoption (UNCTAD, 2001; Daniel et al., 2002; Daniel, 2003). These include “Manufacturing”, “Services”, “Information Technology”, “Finance, Insurance, and Real Estate”, “Wholesale and Retail Trade”, and “Others” (agriculture, communication, utility services). Equal sample sizes (fifty) of firms were selected for each sector. There are unequal numbers of SMEs in these six sectors in Malaysia. Stratified sampling with probabilities not proportional to stratum size (Dawson, 1998) was used to enable comparisons between sectors. While such an approach could restrict the generalisation of the results, it allows for a focus on the issues in industries where e-Business is rapidly becoming institutionalised. Three hundred questionnaires were emailed across these six industries. Overall, 208 respondents returned the questionnaire for a response rate of 69.3%. Sample sizes and response rates are reported in Table 2

<table>
<thead>
<tr>
<th>Sample industries</th>
<th>Respondents</th>
</tr>
</thead>
<tbody>
<tr>
<td>Manufacturing</td>
<td>30</td>
</tr>
<tr>
<td>Services</td>
<td>28</td>
</tr>
<tr>
<td>IT</td>
<td>43</td>
</tr>
<tr>
<td>Finance, Insurance and Real Estate</td>
<td>35</td>
</tr>
<tr>
<td>Wholesale and Retail Trade</td>
<td>32</td>
</tr>
<tr>
<td>Others</td>
<td>40</td>
</tr>
<tr>
<td>Total Respondent</td>
<td>208</td>
</tr>
<tr>
<td>Response Rate (%)</td>
<td>69.3%</td>
</tr>
</tbody>
</table>
3.2 Structural Equation Models

We have employed Structural Equation Modelling (SEM) to test the applicability of our conceptual framework. SEM is a multivariate statistical technique that allows for the simultaneous analysis of the first-order and second-order measurement factors. In our analysis, the first-order factors consist of multi-item measures technological, organizational, and personnel/attitudinal dimensions of each of the basic constructs of supply chain strategy, and e-Business adoption. Supply chain strategy, and e-Business adoption constructs are second-order factors composed of the first-order ones. The dependent measure of business performance is also conceptualized as a “factor of factors” including financial, efficiency, and coordination factors, each of which is composed of multiple items.

The final model provides excellent fit to the data: \( \chi^2 \) of 588.80, \( df = 393 \) with 72 parameters; \( \chi^2/df = 1.50 \); CFI = 0.96; GFI = 0.85; RMSEA = 0.04; TLI = 0.95. This model fit indices fall in an acceptable range (> 0.90) and the RMSEA was less than 0.05. This structural model was nested within the first order model; in that it had been generated by imposing restrictions on, the parameters of the first order model (Figure 2).

Table 3 and Table 4 indicate the hypotheses results for the Malaysian sample. The path coefficients of interest in this model were generated between the independent factors (\( \xi \), exogenous) of e-Business constructs and the dependent factor of business performance (\( \eta \), endogenous). Interestingly, the results suggested that e-Business adoption (H2; \( \gamma = 0.53; c.r. = 4.97 \)) was the strongest predictor of business performance followed by the supply chain strategy construct (H1; \( \gamma = 0.26; c.r. = 2.70 \)). The correlational paths are also of key interest when running this model. Results suggested between correlation between supply chain strategy and e-Business adoption had strong \( \phi \) value of \( \phi = 0.70 \) at significant value of \( t = 1.96 \). The strong correlation was between supply chain strategy and e-Business adoption which confirmed that companies in Malaysia regardless of which sectors they belongs to still treated both of these factors as an important driver for improvement of business performance by treating equally important and they complement each other when a strategy had been formulated.

4 DISCUSSION

The theoretical model confirms that successful e-Supply chain requires supply chain strategy, and e-
Business adoption, which have mutual dependency regardless of geographic and economic differences. For the Malaysian sample (in the context of a developing country), the formation of e-Business adoption is dependent on the implementation of supply chain strategy. This is a critical factor for the Malaysian e-Business development as most of the businesses operate in a larger geographical area.

One explanation of greater relevance of supply chain strategy in the Malaysian sample could be that some of the Malaysian sample surveyed function as a role of contractors to core nations, and may be more focused on primary products. Their success depends on being able to assemble resources and to deliver products on time. The success for the companies operating in core nations may depend more critically on finding new markets for the products. Such an explanation may be viewed as speculation, but the key results are broadly consistent with this sort of a “world systems” view. Results also suggest that the operational differences in managing a global trade and distribution chain are more prominent than any cultural differences in explaining the (limited) differences in the surveyed samples.

The results suggested that companies must pay attention to their technological, organisational, and human capabilities for improving e-Business performance. These capabilities are critical when firms are planning or at the very initial stage of e-Business adoption, where most processes are at low integration levels and are full of manual work (Hsin and Shaw, 2005). Companies that intend to venture into e-Business need to acknowledge and identify barriers caused by “organisation” dimensions by offering training and knowledge for system integration, standards development, and process automation as well as to overcome possible IT resistance.

Where some previous studies have identified supply chain strategy as key dimensions (Wickramatillake et al., 2007; Koh et al., 2006), our model extends this by measuring the impact of technological, organisational and people related issues with e-Business adoption in order to become a successful e-Business firm. Both e-Businesses and conventional businesses use information technology. Our results suggest, however, that technology plays a much more critical role in the business performance of enterprises that have fully adopted the e-Supply chain model. In non-adopting businesses, the use of technology is positively related to business performance, but only modestly so; and, technology use is not integrated with supply chain strategy. In e-Businesses, the use of technology is a stronger determinant of business performance than supply chain strategy. Furthermore, in e-Supply chain, technology use is strongly articulated with business and supply chain strategies.

Adopting enterprises are not without business performance problems. Our results suggest that for e-Business organisations to be successful, supply chain management need to be given a higher level of strategic importance (Koh et al., 2007). We would argue that successful business collaboration is the result of human interactions, which can be supported by IT, but not to be replaced by IT. This is particularly important in the e-Business context where the traditional business model is usually developed on the backbone of technological infrastructure, and "people" related issues can be easily buried under the overwhelming emphasis on technological details. Technology is not the most critical factor in improving supply chains. To improve in this area, SMEs must consider relevant attitudinal issues as identified by Steven (1989) to allow for e-technology to be accepted and diffused in the e-Supply chain.

5 CONCLUSIONS

Our literature review suggests that the existing e-Supply chain and implementation studies lack theoretical underpinning. This situation is more acute for SMEs as the limited numbers of e-Supply chain models found in the literature are not tested empirically. Systems engineering principles, which focus on the interaction of technology, organisation, and people (TOP), provide a useful conceptual scheme for understanding the business performance of both e-Supply chain firms and others.

We have introduced a structural equation modelling approach, and used it to examine the sources of good business performance for companies adopting e-Supply chain. The multi-item constructs of e-Business adoption and supply chain strategy relate differently to business performance. The measures developed here, and the empirical results can be used as a benchmarking tool for the SMEs who wish to embark on e-Supply chain adoption journey. The study also provides some useful directions for new economy cyber-entrepreneurs, guiding them to give due consideration towards appropriate operational and behavioural factors when considering e-Supply chain adoption.
It would be an added value to expand findings obtained from the quantitative study by conducting qualitative investigations in a case study format. As stated by Patton (1987), “case studies are useful where one needs to understand some particular problems in great depth and identify rich information that can be learned from few exemplars of the phenomenon in question”. The future research could be conducted as a complementary study, to further assess and test the applicability of the e-Supply chain factors of e-Business adoption and to identify and investigate any potential benefits, obstacles or emerging themes associated with it. Several organisations (minimum three organisations from each industry) that expressed their interests and met the criteria from both samples could be contacted for face-to-face interview. It is hoped that this combination of quantitative and qualitative study will further support and verify the applicability and robustness of the proposed conceptual model.

REFERENCES


A COMPARISON OF WEB SITE ADOPTION IN SMALL AND LARGE PORTUGUESE FIRMS

Tiago Oliveira and Maria F. O. Martins
Instituto Superior de Estatística e Gestão Informação, Universidade Nova de Lisboa
Campus de Campolide, 1070-312 Lisboa, Portugal
toliveira@isegi.unl.pt, mrfom@isegi.unl.pt

Keywords: Web site, adoption, small firms, large firms, information technology.

Abstract: This study compares the impact of different Technology-Organization-Environment (TOE) factors on the web site adoption decision in small and large firms. A survey that was undertaken by the National Institute of Statistics on the use of Information Technologies (IT) by firms in Portugal was used as the empirical basis for this study. We found significant differences in the factors that determined web site adoption decision in small and large firms. While large firms are mainly influenced by organizational and environmental factors, small firms are also concerned about the technological context. Moreover, the results of our study suggested that, for Portuguese firms, the only factor that is equally important as web site facilitator is competitive pressure.

1 INTRODUCTION

New IT, such as Internet enables firms to do businesses in a different way (Porter, 2001). In order to strengthen the potential of the Internet, firms are establishing their presence on the Web: in 2005, the overall percentage of enterprises in the EU with a web site is 61%, but notably higher for larger firms (90%) than for small firms (56%). Significantly differences also exist between Member States: while the leader countries, Sweden and Denmark, are already reaching the saturation level for large firms (97%), countries like Portugal (75%) and Latvia (65%) are far away from this adoption level. For small firms, this difference is greater: the web site adoption level is 80% for Sweden compared with the 33% level for Portugal (Eurostat, 2006). Do Portuguese small firm managers realize the strategic value of owning a web-site in the same manner as large firm managers? Or have they encountered specific barriers to its implementation? Some studies have been done to understand the differences in IT adoption among European Countries (Zhu et al., 2003) and much research attempted to comprehend the relationship between firms size and IT adoption decision (Lee and Xia, 2006). Some authors (Grandon and Pearson, 2004, Premkumar, 2003) suggested that the research findings on large businesses cannot be generalized to small and medium-sized enterprises (SMEs) because of the unique characteristics of SMEs as for example the lack of business and IT strategy, limited access to capital resources and poor information skills. While there exist an interesting and growing literature addressing the determinants of IT adoption in the specific context of SMEs (Harindranath et al., 2008, Parker and Castleman, 2007,) and a limited research for microfirms (Clayton, 2000), only a reduced number of studies (Daniel and Grimshaw, 2002) attempt to compare directly the approaches of small and large firms to this new domain. Our work seeks to fill this gap in the literature, by analysing the relative importance of the factors that enable or inhibit web site adoption by small firms compared with large firms. The two main purposes of this study are the following:

- To examine the importance of technology-organisational-environmental (TOE) related factors as fundamental determinants of web site adoption;
- To analyze if the relative importance of such factors is different for small and large firms.

To achieve these research objectives we used a rich data set of 637 large firms and 3155 small firms that are representative of Portuguese economy. The understanding of the determinants of web site adoption, at firm level, may be a useful tool in
addressing the right type of policy measures to stimulate the use of internet business solutions, with the aim of enhancing the competitiveness and productivity of Portuguese firms (Bertschek et al., 2006, Black and Lynch, 2001, Bresnahan et al., 2002, Brynjolfsson and Hitt, 2000, Dedrick et al., 2003, Konings and Roodhooft, 2002, Martins and Raposo, 2005, Zhu and Kraemer, 2002). This is particularly needed in the case of Portugal which, for several reasons, has been suffering from a serious lack of competitiveness in comparison to other industrialized economies. Our work has two important contributions: the first is related to the very limited research on comparing the determinants of IT adoption in small and large firms. Secondly, we present useful results for Portugal where there are few published studies on the subject (Parker and Castleman, 2007). The next section presents the theoretical framework based on TOE approach. Then, the proposed hypotheses are tested using an econometric model. Finally, we present major findings and conclusions.

2 THEORETICAL FRAMEWORK AND CONCEPTUAL MODEL

In this study we used the TOE framework, developed by Tornatzky and Fleisher (1990) and applied in many empirical studies related to IT innovations. The TOE model identifies three aspects that influence the adoption and implementation of technical innovations by firms: technological characteristics including factors related to internal and external technologies of firms; organizational factors relating to firm size and scope, characteristics of the managerial structure of the firm, quality of human resources; and environmental factors that incorporate industry competitiveness features. This theoretical background is the one used by Iacovou et al. (1995), Kuan and Chau (2001) and Premkumar and Ramamurthy (1995) to explain electronic data Interchange (EDI) adoption and by Thong (1999) to explain information system (IS) adoption and Hong and Zhu (2006) to explain e-commerce adoption. Empirical findings from these studies confirmed that TOE methodology is a valuable framework to understand the IT adoption decision. In accordance with TOE theory, we developed in the next subsection a conceptual framework for web site adoption (see Figure 1).

2.1 Technology Context

Technology readiness can be defined as technology infrastructure and IT human resources. Technology readiness “is reflected not only by physical assets, but also by human resources that are complementary to physical assets” (Mata et al., 1995). Technology infrastructure establishes a platform on which internet technologies can be built; IT human resources provide the knowledge and skills to develop web applications (Zhu and Kraemer, 2005). Theoretical assertions on the impact of Technology readiness on IT adoption are supported by several empirical studies, based on data sets representative of all sizes of firms (Hong and Zhu, 2006, Zhu et al., 2003, Zhu et al., 2006). These results where also confirmed within the specific context of SMEs (Al-Qirim, 2007, Dholakia and Kshetri, 2004, Kuan and Chau, 2001, Mehrzens et al., 2001). Therefore, in general we expected that firms with greater technology readiness are in a better position to adopt web sites. However, as suggested by others authors (Daniel and Grimshaw, 2002, Parker and Castleman,
2007, Premkumar, 2003), this factor will probably affect in a different way small and large firms.

**H1: The level of technology readiness is positively associated with web site adoption but the impact will vary between large and small firms**

Before the internet, firms had been using technologies to support business activities along their value chain, but many were "islands of automation"—they lacked integration across applications (Hong and Zhu, 2006). The characteristics of the internet may help eradicate the incompatibilities and rigidities of legacy information systems (IS) and accomplish technology integration among various applications and databases. Evidence from the literature suggests that integrated technologies may enhance firm performance by reducing cycle time, improving customer service, and lowering procurement costs (Barua et al., 2004). We define technology integration as the systems for managing orders that are automatically linked with other IT systems of the firm. This type of factor where also identified by Al-Qirim (2007) for the specific case of SMEs. Therefore, we expect firms with a higher level of technology integration to be those who adopt web sites sooner. However, probably there will be significantly differences between small and large firms (Daniel and Grimshaw, 2002). These reflections lead to the following hypothesis:

**H2: The level of technology integration is positively associated with web site adoption, but the impact will vary between small and large firms.**

The lack of security may slow down technological progress. For example, for Portugal in 2002 this was the greatest barrier to internet use (Martins and Oliveira, 2005) and in China it is one of the most important barriers to the adoption of e-commerce (Tan and Ouyang, 2004). We expect firms with a higher level of internal security applications to be more probable web site adopters. Within this context, there is no empirical evidence suggesting a same behaviour between small and large firms. Therefore we stipulate the following:

**H3: Internal security applications are positively associated with web site adoption, but the impact will probably vary between small and large firms.**

### 2.2 Organization Context

Empirical studies consistently found that perceived benefits have a significant impact in IT adoption. This result is validated for medium to large firms (Beatty et al., 2001), for SMEs (Iacovou et al., 1995, Kuan and Chau, 2001) and for all size firms (Gibbs and Kraemer, 2004). However, as suggested by Daniel and Grimshaw (2002) small firms and large firms perceived these benefits in a different. We examine perceived benefits of electronic correspondence and we postulate that:

**H4: Perceived benefits of electronic correspondence is positively related with web adoption, but the impact will vary between small and large firms.**

The presence of skilled labour in a firm increases its ability to absorb and make use of an IT innovation, and therefore is an important determinant of IT diffusion (Caselli and Coleman, 2001, Hollenstein, 2004, Kiiski and Pohjola, 2002). Since the successful implementation of new IT usually requires complex skills, we expect firms with more IT training programs to be more likely to adopt web site. However, there will probably be differences between firms due to the limited IT budgets of small firms. We postulate the following:

**H5: IT training programs are positively associated with web site adoption, but the impact will vary between small and large firms.**

The fact that workers can have access to the IT system from outside of the firm reveals that the organisation is prepared to integrate its technologies. However, this factor is expected to influence in a different way small firms, where the number of employees is small and their presence at the place of work is more important than for large firms. We postulate that:

**H6: The level of access to the IT system from outside of the firm is positively associated with web site adoption, but the impact will vary between small and large firms.**

Regulatory environment has been acknowledged as a critical factor influencing innovation diffusion (Zhu et al., 2003, Zhu et al., 2004, Zhu et al., 2006). Firms often refer inadequate legal protection for online business activities, unclear business laws, and security and privacy as concerns in using web technologies (Kraemer et al., 2006). We postulate that for small firms, this concern will probably be different from their large counterparts.

**H7: The presence of internet and e-mail norms is positively associated with web site adoption, but the impact will vary between small and large firms.**


2.3 Environmental Context

Empirical evidence suggests that competitive pressure is a powerful driver of IT adoption and diffusion (Gibbs and Kraemer, 2004, Hollenstein, 2004, Zhu et al., 2004) and this fact is also verified in small business research (Al-Qirim, 2007, Dholakia and Kshetri, 2004, Grandon and Pearson, 2004, Iacovou et al., 1995, Kuan and Chau, 2001). Therefore, we expect the probability of adopting a web site to be positively influenced by the proportion of web site adopters in the industry or sector to which the specific firm is affiliated. However, some studies suggested that competitive pressure will be more significant in causing small firms to adopt an IT than for larger firms, since they need to protect their competitive position (Daniel and Grimshaw, 2002). Therefore, we assume that:

\[ H_8: \text{The level of web site competitive pressure is positively associated with web site, but the impact will vary between small and large firms.} \]

2.4 Controls

We control, as usual, for industry or economic sector effects. We used a dummy variable to control for data variation that would not be captured by the explanatory variables mentioned before.

3 DATA AND METHODOLOGY

3.1 Data

The data used in this study were provided by National Institute of Statistics (INE) and result from the survey On the use of Communication and Information Technologies in Firms (Iutice) in 2006. In our study we defined that small firms have less than 50 employees and large firms have more than 250 employees. Our sample consists on 3155 small and 637 large firms and is representative of the Portuguese private sector excluding the financial one.

3.2 Methodology

We estimated the following Probit Model:

\[ P(y=1|x) = \Phi(x\beta) \]  (1)

Where \( y = 1 \) if firm decided to adopt a web site, and zero otherwise, \( x \) is the vector of explanatory variables, \( \beta \) the vector of unknown parameters to be estimated, and \( \Phi(.) \) is the standard normal cumulative distribution. To analyse and compare the influence of each factor on the probability of being a web site adopter, we need to compute the marginal effect of \( x_j \). This effect is obtained, for the continuous variables, using the formula given by:

\[ \frac{\partial P(y=1|x)}{\partial x_j} = \phi(x\beta) \beta_j \]  (2)

For the binary explanatory variables it is given by:

\[ \Delta P(y=1|x)/\Delta x_j = \Phi(x\beta|x_j=1) - \Phi(x\beta|x_j=0) \]  (3)

where \( \phi(.) \) is the density standard normal distribution.

The vector of explanatory variables (\( x \)) includes:

**A technology readiness (TR)** index that was built by aggregating 8 items on technologies used by the firm (on a yes/no scale): computers, e-mail, intranet, extranet, own networks that are not the internet (own exclusive networks), wired local area network (LAN), wireless LAN, wide area network (WAN), and one item standing for existence of IT specific skills in the firm (on a yes/no scale) (Zhu et al., 2004). The first 8 items represent the penetration of traditional information technologies, which formed the technological infrastructure (Kwon and Zmud, 1987). The last item represents IT human resources (Mata et al., 1995). To aggregate the items we used multiple correspondence analyses (MCA). The MCA is a method of “multidimensional exploratory statistic” that is used to reduce the dimension when the variables are binary. For more details see (Johnson and Wichern, 1998). The first dimension explains 50% of inertia. In the negative side of the first axis we have variables that represent firms that do not use IT infrastructures and do not have workers with IT skills. On the positive side we have the variables that represent the use of infrastructures and workers with IT skills. Cronbach’s \( \alpha \), the most widely used measure for assessing reliability (Chau, 1999), is equal to 0.8761, indicating adequate reliability. Reliability measures the degree to which items are free from random error, and therefore yield consistent results.

**Technology integration (TI)** was measured by the number of IT systems for managing orders that are automatically linked with other IT systems of the firm (see appendix). The variable ranges from 0 to 5. This variable reflects how well the IT systems are connected on a common platform.

**Internal security applications (ISA)** was measured by the numbers of the use of internal security
applications in the firms (see appendix). The variable range from 0 to 6.

Perceived benefits of electronic correspondence (PBEC) was measured by the shift from traditional postal mail to electronic correspondence as the main standard for business communication, in the last 5 years (on a yes/no scale).

IT training programs (ITTP) is also a binary variable (yes/no) related to the existence of professional training in computer/informatics, available to workers in the firm.

Access to the IT system of the firm (AITSF) was measured by the number of places from which workers access the firm's information system (see appendix). The variable ranges from 0 to 4.

Internet and e-mail norms (IEN) was measured by whether firms have defined norms about internet and e-mail (on a yes/no scale).

Web site competitive pressure (WEBP) is computed as the percentage of firms in each of the 9 industries that had already adopted a web site two years before the time of the survey, i.e. in 2004. As in Zhu et al. (2003) the rationality underlying our model is that an observation of the firm on the adoption behaviour of its competitors influences its own adoption decision.

Services (SER) is a binary variable (yes/no) equal one if firm belong to the service sector.

4 ESTIMATION RESULTS

The web site adoption model is estimated using maximum likelihood. The estimation results for small and large firms are presented in Table 1. Goodness-of-fit is assessed in three ways. First, we used log likelihood test, which reveals that our models are globally statistic significant. Secondly the discrimination power of the model is evaluated using the area under the receiver operating characteristic (ROC) curve, which is equal to 90.9% and 78% for small and large firms, respectively. Finally, the R² shows that the percentage explained by the model is 41.9% for small firms and 15.7% for large firms. The three statistical procedures reveal a substantive model fit, a satisfactory discriminating power and there is evidence to accept an overall significance of the model.

Hypotheses H1-H9 were tested analysing the sign, the magnitude, the statistical significance of the coefficients and the marginal effects. As can be seen from Table 1, for small firms, the estimation results suggested that all the coefficients have the expected signs and the only independent variable that is not statistically significant is the access to the IT system of the firm (AITSF). We can identify seven relevant drivers of web site adoption for small firms: technology readiness (TR), technology integration (TI) and internal security application (ISA) reflecting the technological context; perceived benefits of electronic correspondence (PBEC), IT training programs (ITTP) and internet and e-mail norms (IEN), representing the organization context; web site competitive pressure (WEBP), concerning the environmental context. For large firms, we identify four significant factors influencing web site adoption decision: technology readiness (TR), IT training programs (ITTP), access to the IT system of firms (AITSF) and web site competitive pressure (WEBP). In both cases, as expected, the economic sector is a relevant factor (SER).

Table 1: Estimated coefficients for web site adoption model.

<table>
<thead>
<tr>
<th></th>
<th>Small firms</th>
<th>Large firms</th>
</tr>
</thead>
<tbody>
<tr>
<td>Technological context</td>
<td></td>
<td></td>
</tr>
<tr>
<td>- TR</td>
<td>1.044***</td>
<td>0.346*</td>
</tr>
<tr>
<td>- TI</td>
<td>0.060***</td>
<td>-0.028</td>
</tr>
<tr>
<td>- ISA</td>
<td>0.170***</td>
<td>0.038</td>
</tr>
<tr>
<td>Organizational context</td>
<td></td>
<td></td>
</tr>
<tr>
<td>- PBEC</td>
<td>0.293***</td>
<td>-0.039</td>
</tr>
<tr>
<td>- ITTP</td>
<td>0.235***</td>
<td>0.644***</td>
</tr>
<tr>
<td>- AITSF</td>
<td>0.044</td>
<td>0.278***</td>
</tr>
<tr>
<td>- IEN</td>
<td>0.379***</td>
<td>0.165</td>
</tr>
<tr>
<td>Environmental context</td>
<td></td>
<td></td>
</tr>
<tr>
<td>- WEBP</td>
<td>0.011***</td>
<td>0.017***</td>
</tr>
<tr>
<td>Controls</td>
<td></td>
<td></td>
</tr>
<tr>
<td>- SER</td>
<td>0.185***</td>
<td>0.306**</td>
</tr>
<tr>
<td>Constant</td>
<td>-1.742***</td>
<td>-1.041***</td>
</tr>
<tr>
<td>Sample size</td>
<td>3155</td>
<td>637</td>
</tr>
<tr>
<td>LL</td>
<td>-1038.5</td>
<td>-223.3</td>
</tr>
<tr>
<td>R²</td>
<td>0.419</td>
<td>0.157</td>
</tr>
<tr>
<td>AUC</td>
<td>0.909</td>
<td>0.779</td>
</tr>
</tbody>
</table>

Note: * p-value<0.10; ** p-value<0.05; *** p-value<0.01.

The estimated marginal effects for the determinants of web site adoption model, for small and large firms, are reported in Table 2. Their comparison reveals that, as expected, most of the marginal effects vary between small and large firms. The exception is the web site competitive pressure impact that is the same for small and large firms. Therefore hypotheses H1-H7 are validated and H8 is not confirmed.

There are three additional aspects to be noted here. Firstly, the technological context is much more relevant for small firms than for large firms. Secondly, within organizational context, perceived benefits and internet e-mail norms are more important to determine web site adoption for small
firms than for their larger counterparts. Finally, the access to the IT system of the firm is relevant only for large firms. As a whole, our results are in accordance with those reported in studies comparing IT adoption in large and small firms (Daniel and Grimshaw, 2002). However, the limited number of research in this specific domain difficult the generalization of the results.

Table 2: Estimated marginal effects for web site adoption model.

<table>
<thead>
<tr>
<th></th>
<th>Small firms</th>
<th>Large firms</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Technological context</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TR</td>
<td>0.252***</td>
<td>0.064*</td>
</tr>
<tr>
<td>TI</td>
<td>0.017***</td>
<td>-0.005</td>
</tr>
<tr>
<td>ISA</td>
<td>0.041***</td>
<td>0.007</td>
</tr>
<tr>
<td><strong>Organizational context</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PBEC</td>
<td>0.079***</td>
<td>-0.007</td>
</tr>
<tr>
<td>ITTP</td>
<td>0.061***</td>
<td>0.144***</td>
</tr>
<tr>
<td>AITSF</td>
<td>0.011</td>
<td>0.051***</td>
</tr>
<tr>
<td>IEN</td>
<td>0.100***</td>
<td>0.032</td>
</tr>
<tr>
<td><strong>Environmental context</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>WEBP</td>
<td>0.003***</td>
<td>0.003***</td>
</tr>
<tr>
<td><strong>Controls</strong></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SER</td>
<td>0.044***</td>
<td>0.056**</td>
</tr>
</tbody>
</table>

Note: * p-value<0.10; ** p-value<0.05; *** p-value<0.01.

5 CONCLUSIONS

Within the context of an increased use of Internet Business Solutions, such as web sites, this study fills a gap in the literature by comparing the relative importance of the factors influencing the adoption of web sites for small and large firms. The theoretical framework incorporates most of the facilitators and inhibitor factors identified in other studies. The research model evaluates, for small and large firms, the impact of three technological factors, four organizational factors and one environmental factor on the web site adoption decision. Using a representative sample of Portuguese small and large firms, the estimation results for this comparative study reveal that the important determinants of web site adoption decision vary with size of a firm. Other studies in this domain (Daniel and Grimshaw, 2002, Premkumar, 2003) also suggested that the problems, opportunities, and management issues encountered by small business in the IT area are different from those faced by their larger counterparts. However, our study provides a more in depth analysis since it identifies those factors that more or less relevant for large/small firms and quantifies its impact on web site adoption decision. These findings have practical implications for managers and policy makers. Firstly, policy makers should be conscious that the motivations towards the IT adoption are different for small and large firms. Therefore, government initiatives, such as the Technological Plan, for Portugal, must be different for small and large firms, namely those related to procurement incentives. Secondly, managers should be aware that technology readiness constitutes both physical infrastructure and intangible knowledge such as IT skills. This urges top leaders (mainly in small firms) to foster managerial skills and human resources that possess knowledge of these new information technologies. Therefore, there is a business opportunity for IT firms to establish the service that support the small size firms in the technological context. In our opinion this is particularly important in Portugal given the relative importance of small businesses in the economy (Vicente and Martins, 2008). Finally, our study sought to help firms become more effective in moving from a traditional channel to the internet by identifying the profile of early web site adopters.

As in most empirical studies, our work is limited in several ways. The cross-sectional nature of this study does not allow knowing how this relationship will change over time. To solve this limitation the future research should involve panel data. Another limitation of our work is that it only investigates web site adoption decision. To provide a more balanced view of firms’ IT adoption decision, other Internet Business Solutions, such as e-commerce should also be examined.

ACKNOWLEDGEMENTS

We would like to acknowledge the National Institute of Statistics (INE) for providing us with the data.

REFERENCES


APPENDIX

Technological integration
Did your firm’s IT systems for managing orders link automatically with any of the following IT systems during January 2006? (Yes No)

a) Internal system for re-ordering replacement supplies
b) Invoicing and payment systems
c) Your system for managing production, logistics or service operations
d) Your suppliers’ business systems (for suppliers outside your firm group)
e) Your customers’ business systems (for customers outside your firm group)

Internal security applications

Did your firm use the following internal security applications, during January 2006? (Yes No)

a) Virus checking or protection software
b) Firewalls (software or hardware)
c) Secure servers (support secured protocols such as http)
d) Off-site data backup
e) Subscription of a security service (e.g. antivirus or network intrusion alert)
f) Anti-spam filters (unsolicited e-mails)

Access to the IT system of the firm
Did any of those people access the firm’s computer system from the following places during January 2006? (Yes No)

a) From home
b) From customers or other external business partners’ premises
c) From other geographically dispersed locations of the same firm or firm group
d) During business travels, e.g. from the hotel, airport etc.
PROTOCOL OF AUTHENTICITY TO PROVIDE LEGAL SECURITY IN E-CONTRACTS

A Prototype

João Fábio de Oliveira, Cinthia O. de A. Freitas and Altair Santin
Pontifical Catholic University of Parana - PUCPR
R. Imaculada Conceição, 1155, Prado Velho, 80215-901, Curitiba - PR, Brazil
joao.fabio@gvt.com.br, cinthia@ppgia.pucpr.br, santin@ppgia.pucpr.br

Keywords: Law and Internet, Consumer Protection, Authenticity.

Abstract: This paper discusses the security problems on the contracts deals over the Internet and describes a protocol of authenticity that will keep audit trails from the activities during the Web hiring. These audit trails will be stored digitally, either on the side of the provider or on the consumer. It is understood that it is legal obligation of the provider to logs and ensure the integrity of the data related to the operations on the Internet-based commerce. because in situations of dispute can occur a reversal of the burden of proof. Thus, this prototype ensures confidence on the e-contracts, logging relevant information that help in the identification of the parties, using a Plug-in software installed on the e-commerce provider and in the consumer machine, executable in the Web. Moreover, it is important to remind about the security in contracts agreements over the Internet is an essential feature because it allows the consumer a guarantee of contract award, since it maintains the integrity of the document, and also can be presented as evidence to the Judiciary, helping in litigation and satisfying the premises of the legal acceptance of digital documents.

1 INTRODUCTION

The increased use of Internet in people's daily lives is already a reality as a basic and essential tool in day-to-day of the societies, for things like paying bills, consulting phone catalogs and maps, relationship between people, electronic messaging and even to buy objects and consumer services.

According to the Center for Studies of Information and Communications Technology (Brazilian Internet Management Committee, 2006), about 14.49% of Brazilian homes already have Internet access, this information are from the last poll done in 2006, which represents a 1.56% growth over the 12.93% of 2005. This growing universe of users and potential consumers of online products represents a great concern from the technical and legal point of view, due to the yearly increase in the number of problems that need to be handled by each of the related sciences.

On the other hand, Internet brings some information security concerns, since the documentation is no longer kept in physical form, such as paper, but rather stored electronically through digital means. While the Internet simplifies commercial operations that take place in the digital environment, it introduces a restricting factor and leads us to a universe of studies on the security, trustworthy, confidentiality, integrity, and authenticity aspects. As well as the legality of such operations given the doubtful facts questionable by any of the parties involved in the transaction (e-contract).

In face of this, there is a search for technical solutions for security infrastructure for e-business including the security of the information exchanged and the storage of this information, especially when it comes to making it trusted as far as accuracy of the content stored. In this sense, the cryptography and digital signing methods have contributed for the security of online transactions: e-contract, e-commerce, e-business (Behrens, 2007) (Garfinkel, 1997).

Thus, this paper discusses our prototype for authentication protocol based on the consumer-provider relationship on the contracts over the Internet, since this protocol defines technical and traceable parameters of the transaction, and provides legal security in e-contracts, both in the provider's server and in the consumer's equipment. We discuss our prototype for demonstrating the practical feasibility of the proposed protocol. This protocol
must be used along with the web server of the provider and the Web browser of the consumer, and is been developed as a Web Plug-in.

This paper is divided into five sections. Section 2 highlights the security issues on the web. Section 3 summarizes the e-contracts since its definition and the relevant aspects for the protocol of authenticity, which provides legal security in e-contracts. The proposed protocol is presented in Section 4. Some final considerations are given in Section 5.

2 SECURITY ISSUES ON THE WEB

There are different aspects of security issues on the web, such as physical security, personal security, operations security, communications security, and network security. Normally, the systems are concerned about the application of the protection in e-anything or e-initiative, or specifically e-business, taking into account the state of the art technological infrastructure. The main goal of security for e-business applications is to protect networks and their applications against attacks, ensuring information availability, confidentiality, integrity, auditing, authorization, and authentication (Huang et al., 2008) (Meier, 2003). This paper goes beyond to provide a secure environment. We are here considering legal security of e-contracts.

A secure environment is ensured by combination of different factors, such as data security, networking security, policy, and management of information system security (ISS) (Akhter & Kaya, 2008). Therefore, we need discuss the basis of the information transferred during transactions over the Internet.

The starting point of the information transferred on Internet is the TCP/IP (Transmission Control Protocol/Internet Protocol) protocol. This protocol, in its version 4, has been consolidated for using in Internet (Comer, 1991). This protocol does not include security mechanisms for information transfer, leaving this for the applications which are developed for the end users, which is the reason for the concern about the related criteria to the protection of the content in transit. This means that the transferred information between two different places on Internet, regardless of its physical location, can be captured by a protocol analyser (WireShark, CommView, Ghost). So the desired information can be visualised.

The TCP/IP protocol specification has a conceptual segmentation into five layers, as shown in Table 1 (Comer, 1991). Each layer solves a set of problems involving the transmission of data, and provides a well-defined service to the upper layer protocols based on using services from some lower layers. Upper layers are logically closer to the user and deal with more abstract data, relying on lower layer protocols to translate data into forms that can be physically transmitted. Layer 5, nominated as application layer, specifies and implements the software which interact with the end users. It is at this level that all concerns related to the information security should be implemented, in other words, the applications at the user level must include treatment mechanisms that are considered secure enough to, on one hand, give the end user the assurance that his network transaction is secure, free of risks of content modifications, and on the other hand, provide verified technical conditions of mechanisms considered secure, such as the use of cryptography algorithms at the applications level (Schneier, 1996).

<table>
<thead>
<tr>
<th>Layer</th>
<th>Protocol</th>
</tr>
</thead>
<tbody>
<tr>
<td>5 – Application</td>
<td>HTTP, DNS, SMTP,...</td>
</tr>
<tr>
<td>4 – Transport</td>
<td>TCP, UDP, DCCP...</td>
</tr>
<tr>
<td>3 – Network</td>
<td>IP, IGMP, ICMP...</td>
</tr>
<tr>
<td>2 – Data</td>
<td>Ethernet, Wi-Fi, 802.11(WLAN)...</td>
</tr>
<tr>
<td>1 – Physical</td>
<td>Modem, PLC, optical fiber...</td>
</tr>
</tbody>
</table>

Traffic on the Internet is a client-server software application, which is executed through a direct user interface, known as the web browser. In this environment, several applications are written in the application layer protocol of the TCP/IP known as http (hypertext transfer protocol) (Garfinkel, 1997). Since there are no mechanisms defined in the TCP/IP protocol itself, the application is responsible to define and implement additional security algorithms; minimizing the impact of the vulnerabilities that exist on the internet protocol. Thus, the security issues on the Web are divided into three categories (Garfinkel, 1997):

- **Security of the Web server, the server's activities, and stored data**: guarantee that the information was not modified or distributed without user authorization;
- **Data security in the computers network**: guarantee that the transmission of information between the server and the Web browser has a security level based on well-known standards, such as cryptography or digital signature (Meier, 2003). Cryptography refers to how your application enforces confidentiality and integrity. Digital signature corresponds a
countermeasure techniques that can be used to reduce risk of tampering with data (Meier, 2003);

- **Data security in the user computer:** guarantee that user’s computer is as protected as possible (firewalls, antivirus, anti-spyware, anti-spam, etc.). A typical firewall helps to restrict traffic to HTTP, but the HTTP traffic can contain commands that exploit application vulnerabilities (Meier, 2003). Therefore, taking into consideration the security problems on Internet, these vulnerabilities can be analyzed from the point of view of the contracts deals over the Internet. In this case, it is not necessary provide information security but is primordial provide legal security for e-contracts.

### 3 E-CONTRACT

A contract is a legally binding agreement. Agreement arises as a result of offer and acceptance, but a number of other requirements must be satisfied for an agreement to be legally binding as following (Martin, 2003):

- The parties must have an intention to create legal relations;
- The parties must have capacity to contract;
- The agreement must comply with any formal legal requirements (it may be oral, written, partly oral and partly written, or even implied from conduct);
- The agreement must be in accordance of the law;
- The agreement must not be rendered void either by some common-law or statutory rule or by some inherent defect, such as operative mistake.

However, when such a contract becomes valid, that is, legally binding and enforceable, is regulated by contract law that may differ from country to country, even if subject to International Law. Based on these considerations, an e-contract is a valid contract deals over the Internet. Simplest example is Amazon book store buying. When everything goes right, it is of no major importance when exactly a contract is concluded and according to which law. However, when something goes wrong such as, the product does not delivery, payment is not effected or is wrongly charged, or the product is damaged; then information about the contract conclusion as well as governing law may be crucial. This is the main problem of cross-border e-commerce, but this topic is beyond this paper.

Recognising the value of arbitration as a method of settling disputes arising in the context of international commercial relations and being convinced that the establishment of rules for ad hoc arbitration that are acceptable in countries with different legal, social and economic systems would significantly contribute to the development of harmonious international economic relations; the United Nations Commission on International Trade Law (UNCITRAL) was established by the United Nations General Assembly in 1966 to promote the progressive harmonization and unification of the law of international trade. Brazil not adopts these rules and there are no clear rules or a specific law for litigation situations involving e-commerce.

Chiu et al. (2003) presents that e-contract enforcement can be divided into multiple layers and perspectives, which has not been adequately addressed in the literature. This problem is challenging as it involves monitoring the enactment of business processes in counter parties outside an organization’s boundary. The authors present an architecture for e-contract enforcement with three layers: document layer, business layer, and implementation layer. In the document layer, contracts are composed of different types of clauses. In the business layer, e-contract enforcement activities are defined through the realization of contract clauses as business rules in event-condition-action (ECA) form. In the implementation layer, cross-organizational e-contract enforcement interfaces are implemented with contemporary Enterprise Java Bean and Web services. They present a methodology for the engineering of e-contracts enforcement from a high-level document-view down to the implementation layer based on this architecture, using a supply-chain example. As a result, e-contracts can be seamlessly defined and enforced.

This paper present an authentication protocol to provide legal security in e-contracts implemented in application layer based on document-view. The idea is to be able answering some questions such as: Does the consumer want to buy, is s/he able to buy? When is the contract considered to be concluded? There is no clear rule. To avoid problems and protect the consumer (as provided in the Brazilian Consumer Code) we are developing the protocol to capture and save a set of data from consumer and provider. This set of data can be easily visible and accessible by the consumer and accepted by him/her. Thus, when constituting the contract elements in a virtual environment, the following fundamental
structure items must be established: user-consumer, provider, e-contracts and contract location.

4 PROTOCOL OF AUTHENTICITY

In e-contract, such as in the purchase of consumer goods from e-commerce Web sites, the consumer will be taking advantage all infra-structure of communication defined in the Internet network. When the consumer performs a purchase operation on the providers’ server through her Web browser, she lacks the mechanisms to provide physical and legal evidence of the content accepted during the transaction at a technical level. In other words, there are no effective logs on your computer that store or restore the history of transaction between the consumer and the provider. Therefore, this Section presents the theoretical aspects explored in the proposed protocol and the implementation issues.

4.1 Theoretical Aspects

In practice the hiring conducted on the Internet can be done by e-mail or directly at the Web site (e.g. in clicking a button "I accept"). This implies that specifically in the context of Web communication, the level of information security applied in http protocol guarantees a minimum of security based on mechanisms of encryption. Indeed, the use of SSL (Secure Sockets Layer) in Web applications applies encryption namely, point-to-point, so that the information is transferred ciphered on the Internet (Garfinkel, 1997). The encryption point-to-point means doing a cryptographic channel between the provider and consumer applications layers.

The applications developed for the Internet environment, following the client-server architecture, employ the standard TCP/IP protocol infra-structure, remaining however, vulnerable to the various problems typical to this environment. Specifically in Web communication, when communications between the server and the contractors’ web browser employ the SSL protocol in the cryptography of the data transferred over the network, the risk of fraud in the transferred content is reduced by the cryptography (Garfinkel, 1997). However, in litigation situations it is necessary restore the history of transaction between the consumer and the provider and there is no protocols or tools that capture, store and restore this kind of information, specially providing legal security. In this context, legal security represents the consumer’s guarantees that the transaction hired on the Internet can be presented in Judicial processes. Thus, the proposed protocol of authentication (Plug-in) is organized in two parts:

- **Server:** is installed on the Web server provider and follows the standards of the http protocol and Java-web, and is configured as an extension of the services of the server and offered to the consumer as a Plug-in for web browser of provider;
- **Client:** is installed on the consumer’s computer and follows a specific procedure for initial installation of the Plug-in, considering the consent of the consumer to accept the installation of this software on her computer. The Java programming language was chosen due to its flexibility in selecting of information on Internet environment.

![Figure 1: Plug-in model.](image)

It is possible, therefore, to present an overview of the authentication protocol, in which the consumer possesses in her computer the record of the transactions done in the web site of the providers, in order to extract a report with the information regarding such transaction. On the other hand, the protocol allows to confront the information transferred between both parties, as shown on Figure 1. The request done by the consumer is based on the http protocol, then the Plug-in captures the packages on this level of the protocol as well as other technical information on the TCP/IP level and stores all data in the log files.

4.2 Implementation Issues

Considering the theoretical aspects presented before, the proposed protocol defines that the consumer, when accessing the web site of the providers through her browser, has an icon with information about the authentication protocol available in order to allow her with explicit consent and authorization, to download the Plug-in to her computer.

The protocol determines the capture of consumer information, as shown in Table 2, which takes place during the initial interaction process when the consumer accepts the installation of the Plug-in.
This initial registration makes up the official identification of the consumer.

Table 2: Consumer Information – Initial Registration.

<table>
<thead>
<tr>
<th>Field</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>name_contr</td>
<td>Consumer’s name</td>
</tr>
<tr>
<td>address_contr</td>
<td>Consumer’s address</td>
</tr>
<tr>
<td>phone_home_contr</td>
<td>Consumer’s Home Telephone</td>
</tr>
<tr>
<td>phone_cellular_contr</td>
<td>Consumer’s Mobile Phone</td>
</tr>
<tr>
<td>id_contr</td>
<td>Consumer’s Legal Identification</td>
</tr>
<tr>
<td>type_id</td>
<td>ID, Drivers License, Passport, Working Papers</td>
</tr>
<tr>
<td>tax_id_contr</td>
<td>Consumer’s tax ID</td>
</tr>
</tbody>
</table>

After this initial registration, the algorithm provides verification of the Plug-in installed on the equipment of the provider, installing it when it is not already done. From this point on, the operations performed by the consumer on the web site of the provider keep all information stored in a log file recorded on both sides (provider and consumer), as specified in Table 3.

The log file is encrypted through the Triple DES (Triple Data Encryption Standard) algorithm, developed at IBM and published as a standard in 1977 (Schneier, 1996). Generally TDES uses three different keys and has a key length of 168 bits (3 x 56-bit DES), resulting in a symmetric algorithm. Thus, by being symmetric, the key used for encryption and decryption of the file is the same, making it a fast processing algorithm, considering the log file to be processed and ensuring access to the information.

In this way, the consumer will be able to generate her report directly through her computer, without depending on the information saved in the server of the provider, as is the case with the use of other cryptography algorithms that apply asymmetric keys (which relies on the exchange of public between both parties).

In the Plug-in installed on the computer of the consumer is the function to visualize the log file stored, allowing her extracting the information about the operations accomplished in printed format as evidence of the various Web interactions performed at the web site of the providers. In legal terms, this report will allow the consumer to provide evidence as an official document registered at Notary’s Office (Rezende, 1997).

The capture data follows the structure defined in Figure 2, which is an example of the data capture using the Wireshark software, keeping the reference to the TCP/IP protocol. The information necessary to comply with the fields defined in the authentication protocol are available in the data package captured in the Ethernet frame, as shown in Figure 3. The information captured are stored in the log file in a sequence and logically structured in data files in encrypted text format.

Table 3: e-Contracts Information.

<table>
<thead>
<tr>
<th>Field</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>IP_consumer_Internet</td>
<td>Consumer’s IP number as an Internet user</td>
</tr>
<tr>
<td>IP_consumer_real</td>
<td>Consumer’s IP number (consumer machine)</td>
</tr>
<tr>
<td>Mask_IP_consumer_real</td>
<td>Networking mask (consumer machine)</td>
</tr>
<tr>
<td>Gateway_consumer_real</td>
<td>Default gateway (consumer machine)</td>
</tr>
<tr>
<td>DNS_consumer_real</td>
<td>DNS configured (consumer machine)</td>
</tr>
<tr>
<td>IP_provider_Internet</td>
<td>Provider’s IP number (provider server)</td>
</tr>
<tr>
<td>IP_provider_real</td>
<td>Provider’s IP number (LAN server)</td>
</tr>
<tr>
<td>Mask_IP_provider_real</td>
<td>Networking mask (provider server)</td>
</tr>
<tr>
<td>Gateway_provider_real</td>
<td>Default gateway (provider server)</td>
</tr>
<tr>
<td>DNS_provider_real</td>
<td>DNS configured (provider server)</td>
</tr>
<tr>
<td>date_acess</td>
<td>Date: DD/MM/YYYY</td>
</tr>
<tr>
<td>time_acess_consumer</td>
<td>Hour: HH:MM:SS</td>
</tr>
<tr>
<td>time_acess_provider</td>
<td>Hour: HH:MM:SS</td>
</tr>
<tr>
<td>port_acess_consumer</td>
<td>Consumer’s TCP or UDP port number</td>
</tr>
<tr>
<td>porta_acess_provider</td>
<td>Provider’s TCP or UDP port number</td>
</tr>
<tr>
<td>URL_provider</td>
<td>web address accessed by the consumer</td>
</tr>
<tr>
<td>route_IP</td>
<td>Route between consumer’s IP and provider’s IP</td>
</tr>
</tbody>
</table>

Figure 2: Capturing the Data Package.

The technique for capturing the packages is based on the use of the libpcap/winpcap libraries, which are low level software libraries available for
programming code of development. These libraries come from network traffic information according to the interface used, for example, Ethernet and Wireless WIFI.

| Global Package Package Package Package Package Package Package Package Package Package Package Package Package Package Package Package Package Header Header Data Header Data Header Data Header Data Header Data Header Data |

Figure 3: Format of the Captured Data Package.

This library provides functions that capture packages in the format of the basic network, on which there are the header and the individually separate data. Inside of the TCP/IP protocol in its didactic classification, it is possible to separate the various levels of information (protocol and user information) allowing for the recording of this information to the log file (Comer, 1991).

In future work, we plan to conduct case studies to evaluate our protocol taking into consideration the following parameters:

- **Performance**: computing the impact of the protocol in the consumer machine and on the e-commerce provider based on CPU use. The data collecting will be done using the SNMP (Simple Network Management Protocol);
- **Response time**: which is the time a generic system or functional unit takes to react to a given input. In this case, we will verify the response time between the consumer machine and e-commerce provider. The data collecting will be done using the ICMP (Internet Control Message Protocol) considering the echo-request and echo-reply facilities.

This evaluations can demonstrate potential scenarios that may benefit from this research.

5 FINAL CONSIDERATIONS

In litigation situations involving e-commerce, the greatest difficulty is in verifying proof of the contract established. This difficulty arises due to the fact that the relations are no longer necessarily face-to-face, therefore requiring the use of additional mechanisms to carry out these contracts as well as computer tools that allow the registration and evidence that the contract was made. Thus, the authentication protocol proposed in this article allows both interested parties, consumer and provider, to keep registry logs with information about the dealing contract. This audit trail is composed of a variety of information, such as the IP record of the equipment involved in the transaction. Therefore, the consumer can issue reports on the access to the providers, as well as the contracts over the Internet. In this situation, it is important that the consumer can restore, from her own computer, data and information on the litigated web site. Or, furthermore, the consumer should be in condition to validate the information given by the provider. Thus, the use of the authentication protocol will be, in fact, mapping the operation accomplished through the Internet by means of the log file, with a full and secure record of the main elements of the e-contract.

ACKNOWLEDGEMENTS

The authors wish to thank Dr. Antonio Carlos Efing (PPGD-PUCPR) which have collaborated to this work. This work has been supported by CNPq (grant 471627/2006-2 and grant 476637/2006-6).

REFERENCES


USING TECHNOLOGY ACCEPTANCE MODEL TO EVALUATE USERS’ ATTITUDE AND INTENTION OF USES

Dauw-Song Zhu
Business Administration & Accounting Department, National Dong Hwa University
No. 1, Sec. 2, Da Hsueh Rd. Shou-Feng Hualien, Taiwan 974
dswu@mail.ndhu.edu.tw

Chih-Te Lin
Business Administration Department, National Dong Hwa University &
Lecturer of Food and Beverage Management Department, Taiwan Hospitality & Tourism College
No. 2, Lane 46, Jhongmei 3rd St, Hualien, Taiwan 970
thomsonhua@gmail.com

Keywords: Online game, technology acceptance model (TAM), service quality.

Abstract: Many empirical studies have pointed out that the technology acceptance model (TAM) can be used to explain whether users can accept a new information technology. Therefore, this study has adopted TAM to investigate external factors that affect gamers’ acceptance of online games. In this study, system quality, information quality, and service provider’s characteristics were taken as external variables. It was discovered that in the aspect of perceptions, system quality had positive effects on perceived ease of use and perceived usefulness. Service provider’s characteristics had positive effects on perceived usefulness and perceived trust. Besides, in the relationship between user’s perception and attitude and intention, the research finding was consistent with TAM has suggested; i.e. perceived ease of use had positive effects on perceived usefulness and user’s attitude, and perceived usefulness had positive effect on user’s attitude. Finally, user’s perceived trust and attitude would be positively correlated with user’s intention of use.

1 INTRODUCTION

With the advancement of information technology, popularity of the Internet, and the gradual prevalence of broadband networks, the output value of online games has been increasing in a fast speed. In fact, Internet websites are accessible by consumers from around the globe, and this has contributed to the globalization of markets over the past three decades (Yip, 2000), it is very common for researchers studying the international environment to select one country and use it as a basis for empirical inquiry. Clearly, managers need to better understand the growing use of Internet shopping sites and those consumer characteristics that encourage repeat visits to these sites. One key consumer characteristic is their willingness to use and accept the new technologies. Since researchers need a framework to effectively evaluate new online phenomena, the technology acceptance model (TAM) is applicable in this context (Savitskie, K., Marla B Royne M.B., Persinger, E.S., Grunhagen, M., Witte, C.L., 2007). The TAM model was derived from Fishbein and Ajzen’s (1975) Theory of Reasoned Action and has received considerable support by many researchers (e.g. Jackson, Chow, & Leitch 1997; Venkatesh & Davis 1996). Further, a number of researchers have examined Internet usage via the TAM framework (Gefen & Staub 2000; Venkatesh, & Massey 2003), and the current research use the TAM to better understand individual attitudes toward technology within the international Internet shopping environment.

Moreover, the identified human factors are mainly user perceptions about the utility of this type of technology: perceived ease of use and perceived usefulness (Davis, 1989). Some related studies have also revealed that when one enters the fully immersed state, he will feel totally involved in the activity, his consciousness will be very narrow, his attention is only focused on the activity, and his awareness is partially lost (Webster et al., 1993;
Koufaris, 2002). However, the gaming market is very competitive, and several new games are released every month. Thus, when players have a certain degree of ‘flow’ experiences from a new game, they may quit playing it due to awful external quality of the game. As discussed above, the purpose of this study is to investigate the effects of external factors, such as system quality, information quality, service provider’s characteristics on the perceptions of online gamers, and find out whether gamers’ perceptions have positive effects on their attitude for the game and intention of use, in hope of providing the research result as a reference for online game companies to enhance service quality for customers.

2 LITERATURE REVIEW

2.1 System Quality (SQ)

Website quality, system quality, and game quality are significantly influential to gamer satisfaction. In the I/S success model proposed by DeLone (1992), 12 empirical studies about system quality were investigated, and 18 indicators of system quality were proposed. These indicators include ease of use, usefulness, system accuracy, system flexibility, system reliability, response time, and etc. However, Mckinney, Yoon, and Zahedi (2002) measured system quality of Internet shops with the following indicators, including access, usability, navigation, and interactivity. In Negash (2003) which focused on online customer service systems, it was argued that some of DeLone’s system quality indicators are already outdated. Thus, interactivity and access were proposed as system quality indicators.

2.2 Information Quality (IQ)

As online games will be constantly updated and expanded, game information is very important for online gamers. In the evaluation of information quality, Lee, Strong, Kahn, and Wang (2002) developed 15 indicators to assess the information quality of an organization. These indicators are accessibility, appropriate amount, believability, completeness, concise representation, consistent representation, ease of operation, free of error, interpretability, objectivity, relevancy, reputation, security, timeliness, and understandability. In Mckinney et al. (2002), only 5 indicators, including relevance, timeliness, reliability, and scope, were adopted. Negash et al. (2003) proposed to use informativeness and entertainment as system quality indicators.

2.3 Service Provider’s Characteristics (SPC)

Saeed et al. (2003) conceived that service provider’s characteristics are important because the Internet is a virtual channel which creates more sense of uncertainty to Internet users. Stronger service provider’s characteristics will enhance users’ trust and consumers’ perceptions will not be affected. Saeed et al. further proposed the following indicators to measure service provider’s characteristics, including size, reputation, and participation costs. In Jarvenpaa, Tractinsky, and Vitale (2000), it was proposed that perceived reputation and perceived size of service providers will affect users’ trust for service providers. Among domestic studies, Chen Chuen-Liang (2002) pointed out that online game’s brand image will positively affect users’ intention of use, and brand image can be evaluated by corporate image, word of mouth, popularity, and reputation.

2.4 Technology Acceptance Model

Over the past 10 years, Technology Acceptance Model (TAM) has been empirically proven to be an important explanatory model for personal acceptance or use of new information technology. TAM is a behavior intention model developed based on the Theory of Reasoned Action (TRA) by Davis in 1989. It was intended to simplify TRA and find out an effective behavior model that could be widely applied to explain or predict the factors affecting the use of information technology. In TAM, two definite cognitive beliefs were proposed, namely perceived usefulness and perceived ease of use. The two beliefs determine an individual’s behavior intention for using technology through attitude. It has been clearly pointed out that external variables will directly influence perceived usefulness and perceived ease of use and indirectly affect user’s attitude, intention, and practical use. Based on the behaviors of information system users, Seddon et al. (1997) developed a successful information system model which definitely pointed out that system quality and information quality would respectively affect user’s perceived usefulness and perceived ease of use. Lin and Lu (2000) probed into the behavior intentions of World Wide Web (WWW) users, using information system quality to measure user’s behavior intention for using WWW. It was discovered that system and information quality were positively correlated with user’s cognition, attitude,
and willingness. In a review of studies related to online consumer behaviors, Saeed, Hwang, and Yi (2003) proposed a set of integrative structure, in which system quality, information quality, service quality, and service provider's characteristics are influential to user's perceived ease of use, perceived usefulness, and perceived trust.

In recent years, many empirical studies (David, 1989; Szajna, 1996, Lederer et al., 2000; Lin and Lu, 2000; Moon and Kim, 2001; Hsu and Lu, 2003) have verified that perceived ease of use would affect perceived usefulness, perceived usefulness and perceived ease of use would affect user's attitude, and user’s attitude would further affect user’s intention of use.

3 RESEARCH METHODS

In this study, the questionnaires were mainly distributed to players of a new online game available for public test. The game was selected according to the observation of popular forum topics on a well-known game website “Bahamut” (www.gamer.com.tw) during one week (Apr 27–May 3 2005). The statistic result revealed that the “RF Online” forum was almost always on the top 3. During that time, “RF Online” was available for public test, so it was selected as the research focus. Later, a web-based questionnaire was formed and published on a professional web-based survey website “MY3Q”. The survey link was also posted on the online game forum of BBS at Dong Hwa University, famous game forums "Bahamut" and “Game Base”, and some game-related communities on Yahoo for players to connect to the survey system. The survey period started from May 5 till May 31 2005. A total of 319 valid samples were collected.

3.1 Measurement of Variable

According to Mckinny et al. (2002), system quality was divided into three dimensions, including “access”, “usability”, and “navigation”. Access was defined as the connection response and access speed of the game and the website. Usability refers to the operations of the user interface in the game, and navigation is defined as the operation of the user interface on the website. Aladwani and Palvia (2002) pointed out that security mechanism is also an important element when users evaluate the quality of a website. Thus, this study also incorporated “security” as dimension of system quality and defined it as the level of security of gamers’ personal data and the gaming process.

Based on Mckinny et al. (2002) and Lee et al. (2002), this study proposed 4 dimensions of information quality, namely “relevance”, “timeliness”, “reliability”, and “scope”. Relevance refers to the applicability of the game information provided by the website. Timeliness indicates whether the website can provide latest game information. Reliability is defined as the correctness of the game information, and scope refers to the coverage of the provided game information.

As suggested in Jarvenpaa et al. (2000), service provider’s characteristics included two dimensions, reputation and perceived size. Reputation is defined as the prestige of the firm, and perceived size refers to the scale of the company in this industry. According to TAM and the study of online guys by Hsu and Lu (2003), perceived ease of use was defined as the level of easiness that users feel about the functions of an online game. Based on TAM and Hsu and Lu (2003), perceived usefulness was defined as the level to which users feel that the online game can achieve the gaming objective. In Hsu and Lu (2003), gaming objectives included fun, recreation, messaging, information exchange, making friends, chatting, team work, fantasy, hobby, work, and transaction. This study employed TAM and Hsu and Lu (2003) to define attitude as the level of user's preference for an online game. Based on TAM and Hsu and Lu (2003), this study defined the intention of use as the intensity of user’s intention to use an online game. After all the questionnaires were collected, we analyze the data and verify the hypotheses with structured equation model (SEM). The analysis procedure included two parts, basic analysis and overall model analysis. In the basic analysis, descriptive analysis and reliability analysis would be processed on SPSS 10. In the overall model analysis, confirmatory factor analysis and structural equation model analysis would be performed on Amos 4.0.

4 RESULTS

In the gender distribution among collected samples, male players accounted for 84.9% and female ones only 15%. Players aged between 19-23 took the largest proportion by 37.9%. Players aged between 19-23 took the largest proportion by 37.9%. 45.4% of them had a university or college education background, and 38.8% had a high school or vocational school education background. In terms of occupation, 66.7% of them were students. This sample structure was similar to those observed in other studies of
online games but not in the studies of behaviors on the Internet. As to gender distribution, the ratio of male and female online gamers was 7:3. In the aspect of online shopping, female shoppers outnumbered male ones, by a ratio of 6:4. Other activities, such as the use of portal sites and Internet phone, had a relatively even ratio of male to female users. In the age distribution, if age 24 was viewed as a barrier, the ratio of those above 24 to those under 24 was 7:3.

Table 1: Path coefficients of the proposed model and verification of hypotheses.

<table>
<thead>
<tr>
<th>Variables</th>
<th>Beta</th>
<th>Ha</th>
<th>Result</th>
</tr>
</thead>
<tbody>
<tr>
<td>SQ → perceived ease of use</td>
<td>1.375**</td>
<td>H1</td>
<td>○</td>
</tr>
<tr>
<td>IQ → perceived ease of use</td>
<td>0.303*</td>
<td>H2</td>
<td>○</td>
</tr>
<tr>
<td>SQ → perceived usefulness</td>
<td>NA</td>
<td>H3</td>
<td>×</td>
</tr>
<tr>
<td>IQ → perceived usefulness</td>
<td>NA</td>
<td>H4</td>
<td>×</td>
</tr>
<tr>
<td>SPC → perceived ease of use</td>
<td>0.859**</td>
<td>H5</td>
<td>○</td>
</tr>
<tr>
<td>Perceived ease of use → perceived usefulness</td>
<td>0.424**</td>
<td>H6</td>
<td>○</td>
</tr>
<tr>
<td>Perceived ease of use → user’s attitude</td>
<td>0.363**</td>
<td>H7</td>
<td>○</td>
</tr>
<tr>
<td>Perceived usefulness → user’s attitude</td>
<td>0.314**</td>
<td>H8</td>
<td>○</td>
</tr>
<tr>
<td>Perceived usefulness → intention of use</td>
<td>0.833**</td>
<td>H9</td>
<td>○</td>
</tr>
<tr>
<td>User’s attitude → intention of use</td>
<td>0.168**</td>
<td>H10</td>
<td>○</td>
</tr>
</tbody>
</table>

**p<0.001

However, the ratio was reversed in the aspect of online shopping, where those above 24 took the largest proportion (6:4). In other activities on the Internet, both groups had an even distribution. In terms of occupation, most of the online gamers were students. The ratio of student gamers to non-student gamers was 6:4. But in the aspect of online shopping, those of other occupations took a larger proportion, and students only accounted for 30%. Besides, no significant difference between the two groups was observed in other activities. Finally, in terms of education background, the majority of online gamers received college education. It can be discovered that every dimension had a mean larger than 3, and attitude and intention of use even had 4.26 and 4.28, respectively, indicating that players had a high level of preference for the online game.

In this study, AMOS 4.0 was used to process the structural equation model analysis and verify the proposed causal path model. After the path was established and the samples were applied into the model, it was discovered that many relationships among latent variables were not significant. Thus, the structure was modified to remove the insignificant path between exogenous and endogenous variables.

As to basic goodness of fit, the factor loading of each indicator ranged between 0.5 and 0.95 and reached the level of significance. Besides, there was no negative deviation. Thus, overall, this model was compliant with the standard of goodness of fit. As to the fitness of overall model, the absolute fit measures of the overall theoretic model: $\chi^2$/d.f=1.90, GFI=0.87, RMR=0.04, and RMSEA=0.05, where $\chi^2$/d.f, RMSEA, and RMR reached the ideal level, and GFI also approached the acceptable level of 0.9. In the aspect of incremental fit measures, AGFI=0.84, NFI=0.90, CFI=0.95, NFI and CFI exceeded the ideal level of 0.9., and AGFI was also close to the acceptable level of 0.9. Among parsimonious fit measures, PNFI=0.80 and PGFI=0.72, all of which were compliant with the standard.

The path coefficients in the path model and the results of hypotheses verification were organized in Table 1. In the aspect of the impact of external factors on user’s perceptions, user’s perceived ease of use was affected by system quality and information quality. Thus H1 and H2 were supported. In the aspect of perceived usefulness, the research findings revealed that system quality and information quality had no effect on perceived usefulness, so H3 and H4 were not supported. However, service provider’s characteristics and perceived ease of use had positive effect on perceived usefulness. Thus, H5 and H6 were supported. Perceived ease of use and perceived usefulness had positive influence on user’s attitude. Thus, H7 and H8 were supported. Finally, perceived usefulness and user’s attitude (H9, H10) would have positive influence on intention of use at the same time.

5 CONCLUSIONS

System quality and information quality were
positively related to player’s perceived ease of use. This indicates that when better and more stable systems and correct and rich information are provided to game players, game players will feel perceive more ease of use and the entry obstacle of the game can be reduced.

No significant effect of system quality and information quality on gamer’s perceived usefulness was discovered. Unlike task-oriented information systems, online games are entertainment-oriented information systems. Thus, when players are engaged in online games, they do not necessarily play the game to enhance their game performance or seek higher efficiency but simply kill some time, make friends, get rid of the social bindings. This explains why system quality and information quality were not significantly related to perceived usefulness.

Service provider’s characteristics had positive effects on Perceived usefulness. In online games, it would be time-consuming to accumulate achievements or cultivate relationships, so if the service provider is in a small scale or it does not proper manage with players, the service provider may shut down the game due to improper management.

In this study, TAM was adopted to investigate the factors affecting online game players’ acceptance of games. The research results were consistent with those suggested in previous studies. This shows that if gamers feel that a game is easy to be familiar with and get involved in, they will have more preferences for the game and further increase their intention to carry on using it.

6 MANAGERIAL IMPLICATIONS
Most of the previous studies of TAM focused on the second half of the model, i.e. the effect of user’s perceptions on their attitude and intention. The main contribution of this study is that it probed into the first half of the TAM model to understand the impact of external factors on user’s perceptions.

In the previous studies about evaluation of information systems, external factors were mainly touched upon in the discussion of the impact of information quality and system quality on information systems. However, with the fast development of the Internet, many related online information systems have been derived. Thus, the past evaluation models are no longer applicable to the evaluation of online information systems. In this study, the importance of service provider’s characteristics was empirically proven. Online users cannot directly interact with service distributors or providers, so trust becomes an important element in the evaluation of online information systems, and well-known and large-scale firms can usually lower the level of uncertainty.

The research of online games is seldom across to foreign nations, and no measurement scale for online games is available. This study try to use the scales designed for other information systems, and incorporated the opinions of online gamers to develop an integrated scale for online game systems. The proposed model had compliant validity and reliability. Thus, it can be a reference for further studies.

Besides, in the aspect of brand image, a good management mechanism is helpful for the establishment of a brand. If game service providers can make use of brand advantage, they can attract more users to participate in their games and enhance customer loyalty. Besides, a good brand image can also help promote other games, create popularity, and establish a good reputation. Finally, according to the survey of this research, 54% of the gamers reported to play other games in addition to the selected game. This reveals that there are numerous choices for online games in Taiwan, and the market competition is very fierce. As a result, if game service providers are imprudent in their management, their games may be easily displaced.

7 RESEARCH LIMITATIONS
In this study, a web-based survey was adopted and the reference link was posted on some major forums and discussion boards for game players to participate in the survey at will. Thus, random sampling could not be conducted, and deviation of the samples in the representation of the population might occur and possibly lead to a slight bias of the research results.

REFERENCES


Jackson, C.M., Chow, S., & Leitch, R.A.(1997).Toward an understanding of the behavioral intentions to use an information system. Decision Sciences, 28(2),357-389


SEMANTIC INTEROPERABILITY FOR E-BUSINESS IN THE ISP SERVICE DOMAIN

Jane Hall
Fraunhofer FOKUS, Kaiserin-Augusta-Allee 31, 10859 Berlin, Germany
jane.hall@fokus.fraunhofer.de

Stefanos Koukoulas
Eworx S.A., 66 Jean Moreas Sr, 152 31 Halandri, Athens, Greece
sko@eworx.gr

Keywords: Ontology, semantic interoperability, eBusiness, virtual organisation, ISPs.

Abstract: Enterprise interoperability is a challenging goal that has many facets, one of which being the need for cooperating enterprises to have a precise understanding of the information that they exchange with each other. Internet Service Providers (ISPs) collaborating in a virtual cluster to market, customize and provision services to customers require the same understanding of these services and their features. This paper describes the work being undertaken in the European IST project VISP (IST-FP6-027178) to meet this requirement. ISP services are being classified and formally specified and an ontology is being developed for the ISP service domain.

1 INTRODUCTION

Enterprises wishing to undertake eBusiness by cooperating in a virtual organization must agree on the definitions and meanings of concepts used in their cooperation, and the software infrastructure supporting their eBusiness transactions must be aligned with this understanding so that services can be offered and provisioned for customers dynamically, efficiently and effectively. This paper is concerned with the work being developed in the European IST project VISP (IST-FP6-027178) to support semantic interoperability for a virtual cluster of small ISPs collaborating to market, customize and provision services for business customers.

The structure of the paper is as follows. First the eBusiness context of small ISPs and their collaboration in a virtual cluster is briefly outlined. Reference is then made to the significance of knowledge and semantics for enterprise interoperability in the European research area and to research projects here. Work that has been undertaken in the VISP project on the specification and classification of ISP services is introduced. Then the ontology that has been developed for the services in the ISP domain is presented, including an application of its use in composing services for eBusiness. A brief summary of the work undertaken concludes the paper.

2 ISPS AND VISP

Since the advent of the Internet, the traditional ISP market has been in constant evolution due to the gradual globalization and commoditization of ISP services and to deregulation initiatives aimed at fostering competition. Small ISPs are best at targeting niche markets; they can respond rapidly and provide excellent customer service. However, small ISPs cannot offer the wide range of services now required by increasingly demanding business customers nor offer the geographical coverage that is necessary in a globalized marketplace.

Collaborating in a virtual cluster to undertake eBusiness enables small ISPs to offer a wider range of services, particularly when customizing tailored services that are composed from individual services. Tailoring composite services from the services of other partners in the cluster in order to offer innovative and higher value solutions to customers is indeed one of the principle advantages and reasons...
behind the idea of collaborating in a virtual VISP cluster with other small SMEs.

However, when composing services it is necessary to have a precise understanding of the features of each service element so that the composition of individual services into a service bundle is consistent and the composite service can be validated. In order to meet these requirements, work was undertaken within the VISP project to develop a service decomposition and characterization methodology and to establish an ontology for semantic interoperability within the cluster.

### 3 KNOWLEDGE MANAGEMENT AND ONTOLOGY FOR ENTERPRISE INTEROPERABILITY

The relevance of knowledge management and semantic interoperability for cooperating enterprises is a topic of current significance. One of the four Grand Challenges in the Enterprise Interoperability Research Roadmap (Li et al., 2006) is Knowledge-Oriented Collaboration, now renamed Knowledge-Oriented Collaboration and Semantic Interoperability to emphasize the significance of semantics in enterprise interoperability (Charalabidis et al., 2008). Knowledge-based collaboration allows the sharing of knowledge within virtual organisations to the mutual benefit of all partners. The advances being undertaken in the area of knowledge management and ontology development are intended to be of direct benefit to enterprise collaboration. Those sharing the knowledge need have the same understanding of it, hence the requirement for ontology definitions in this area.

Despite a rapidly changing research area, work already undertaken was studied in the VISP project to establish the state of the art in ontology specification methods and ontology management techniques. Existing approaches were examined and their strengths and weaknesses evaluated in conjunction with the specific requirements of the VISP project for the use of an ontology.

There was an exhaustive state of the art analysis in many of the fields relevant to VISP (DIP, 2004). It developed the DIP Ontology Management Suite, which is an integrated set of tools for efficiently and effectively managing ontologies. VISP was able to input this work into its own state of the art analysis.

The DIP IST project objective was to develop and extend Semantic Web and Web service technologies, and it undertook an exhaustive state of the art analysis in many of the fields relevant to VISP (DIP, 2004). It developed the DIP Ontology Management Suite, which is an integrated set of tools for efficiently and effectively managing ontologies. VISP was able to input this work into its own state of the art analysis.

The work carried out in these and other projects provided generic results that could be evaluated and used in VISP as a basis for developing a specific ontology for services in the ISP domain.

### 4 SPECIFYING ISP SERVICES IN VISP

A VISP cluster can only work efficiently if all partners have the same view not only of what a service is and what it offers but also of relationships and constraints between services, service characteristics and characteristic values. In a federation of independent autonomous entities, there is no centralized control but there must be a common understanding between the partners of services and their properties in order to collaborate dynamically in real time. Properly specifying the properties of the elementary service components is “a key aspect of supporting the proper specification of (composite)
services’ semantics” (Leymann, 2007). The composite service must behave correctly when provisioned for the customer.

A primary objective of the VISP project is to build an ISP service knowledge base. A service has been defined in VISP as representing an atomic set of capabilities or activities that can be provisioned and used alone, or that can be grouped with other services to form a composite service. A set of elementary services for the virtual cluster has been investigated, classified and specified. The result is a knowledge base of ISP services formally represented using standard representations.

The VISP Service Knowledge Base (SKB) is a service catalogue that contains the technical information relating to the individual services that can be used to offer services to customers either standalone or as part of a composite service. The SKB thus provides a common technical understanding of services in the cluster, i.e. a common “language” shared by all partners that ensures consistency within the cluster.

Services described in the SKB are organized in categories and sub-categories. The nine categories comprise access services, bandwidth services, hosting services, application services, network services, security services, support services, and also software supplies and material supplies. Each of the nine categories defined is divided into sub-cATEGORIES. The SKB can be centralized, distributed or replicated at each partner, depending on the functional architecture adopted.

A basic service modelling approach was adopted. A service is specified by a list of characteristics and each characteristic comprises a list of values. The service, each characteristic and each value has an associated description plus other parameters. This approach is very simple, generic and reusable. It does not depend on any particular technical or business environment and is applicable to many fields apart from the ISP/telecom domain.

Each service is thus characterized by a general description, a set of mandatory and optional parameters with their possible values, usage limitations and exclusions together with information on how they can be composed. All services offered by partners in the cluster are described using this specification. This information constitutes the VISP service knowledge base that, as the reference for ISP services in a VISP cluster, is being used to specify the VISP ontology.

5 ONTOLOGY FOR THE ISP DOMAIN

An ontology is being developed in the VISP project for VISP purposes, i.e. to enable partners in the cluster to collaborate in providing tailored services to customers as dynamically and in as automated a manner as possible. An ontology can be regarded as a “set of shared conceptualisations of entities within an application domain” (Wilson et al., 2006). In order to carry out eBusiness dynamically and in close cooperation with other partners in a virtual enterprise, all partners in the cluster need to adopt a common semantic model for their interactions, a shared understanding that can be automated by using a software infrastructure supporting the eBusiness of the ISPs. The VISP ontology thus constitutes one of the innovative results of the project.

Although problems are associated with the development of ontologies (Hepp, 2007), it was felt that the requirement for automating what can be a complex task meant that an ontology was essential for a VISP cluster. Work undertaken on developing an ontology for service components is often of a quite generic nature, for example (Sheth et al., 2006) and (Wang and Xu, 2008). However, the ontology here is being developed for VISP cluster purposes, to support the validation of ISP services composed automatically and dynamically from other services in an eBusiness context. It therefore has a restricted purpose and a limited circle of users and it should not be the bottleneck mentioned by Hepp when ontologies that are created by a small community are intended for much wider use.

The main use of an ontology in VISP is in the area of service composition and tailoring of innovative and often complex customized services. When new services are composed of individual component services, it must be ensured that the features of each service comprising the composite service do not conflict or result in inconsistency in the tailored service. The use of ontology is intended to assist the representation of the complex semantic and the relationships between the services offered by the partners in a VISP cluster. Validation of a service bundle requires a validation of the rules and restrictions among all services that form the service bundle. Semantic information such as restrictions, relationships between services, rules for the values of characteristics, etc. is static. It was therefore decided to model the content of the SKB using an ontology whose instance data are the service descriptions of all VISP services.
An analysis was performed within the VISP project to compare ontology tools and languages and to decide on the approach to take given the VISP requirements. Languages such as Topic Maps (www.isotopicmaps.org/sam/sam-model), RDF Schema specification (www.w3.org/TR/rdf-schema), OWL Web Ontology Language, (www.w3.org/TR/owl-ref/), OWL-S (www.w3.org/Submission/OWL-S) and F-Logic (Kifer et al., 1995) were investigated. Ontology editors were evaluated based on (Gomez-Perez et al., 2002) and work performed in SEKT (SEKT, 2004). Repositories for storing and retrieving ontologies and ontology change management systems were also evaluated.

Based on the evaluation results and the requirements of the VISP cluster in doing eBusiness, Protégé (http://protege.stanford.edu) was selected as the ontology editing tool, providing not only ontology editing capabilities but also constituting ontology library systems and offering various functions for managing, adapting and standardizing groups of ontologies while enabling ontology reuse. No tool fulfilled all the requirements for a reliable, secure, interoperable ontology management environment although most seemed to provide adequate solutions for most of the main issues in the VISP eBusiness environment. Protégé was one of the first editors available, it is open source and supports a variety of plugins and import formats, such as RDF Schema and OWL. It is also widely used; 68.2% of respondents in a survey were using Protégé as their ontology editor (Cardoso, 2007).

The principle requirement for an ontology specification language in VISP is expressive power and OWL was selected here. It supports the RDF data model, is used in the Semantic Web world and was felt to be the most appropriate choice for the eBusiness domain to be modelled in VISP. It too is widely used; 75.9% of respondents in the survey mentioned above are using OWL to develop their ontologies (Cardoso, 2007).

A two-phase methodology was developed for designing, developing and using the ontology-based knowledge management system for the VISP software environment. In the first phase, the application area and use of the ontology were determined. In the second phase, the ontology was built based on the information sources identified in the first phase and the concepts used there. Although the area of tools and technologies enabling the application of ontologies on a business level is not considered mature and much work is still in progress, the use of an ontology in VISP was recommended. Given the increasingly complex requirements of the VISP eBusiness environment and the need for rich, consistent and reusable semantics, ontologies represent an optimal answer to the demand for an intelligent system that operates more closely to the human conceptual level.

5.1 Description of the SKB Ontology

The ontology has been developed in VISP to represent the semantic information included in the VISP domain, i.e. ISP services. This ontology is intended to model the SKB but would also describe the classification of service descriptions, the attributes of service descriptions as well as the rules, constraints and relationships between them.

The service descriptions were classified as presented in section 4 above. The goal of the SKB ontology definition is to support this classification of service descriptions, to include the attributes of service descriptions and the relationships between them as well as to demonstrate the rules and the constraints for these attributes in a formal way that can be used to validate a composite service, or service set as it is termed in VISP.
are not required in the definition of an ontology, the primary purpose of which is to enable the validation of a service set from within the VISP software infrastructure. Such information includes mainly fields that are used to describe static attributes inside the SKB and which are of no special interest inside the SKB ontology, for example, Description, GenericDeploymentInformation, References.

The second view is the description of properties. OWL properties model all relationships between individuals of different classes. Apart from this pure listing of the ontology properties, this model also includes a description of the restrictions that each property of each class should satisfy. This aspect is further discussed in the following section.

Figure 2: List of OWL properties in the SKB ontology.

The third view is the description of individuals. Individuals are instances of classes such as in object-oriented programming languages. Thus the descriptions of classes mentioned above also provide general descriptions for the individuals of these classes.

5.2 Restrictions to be Represented

The use of an ontology cannot only be justified by the need to model the domain of interest in a fairly simple way by using OWL classes and properties that resemble capabilities of UML diagrams and object-oriented design. The need for an ontology has evolved due to the need to define complicated rules and constraints governing the domain of interest (in this case the ISP service domain). The definition of these restrictions is probably the most important part of the SKB ontology in the sense that it differentiates it from a flat database schema such as the SKB itself. Such rules and constraints can be defined in OWL provided that it is possible to express them as logical expressions.

Therefore, the definition of the SKB ontology is a two-step process:

1. First, the “skeleton” ontology model has to be specified. The class hierarchy, the listing of properties and the definition of individuals form the necessary schema elements of this model and will be utilized to describe the domain restrictions in the next step.

2. Then the rules and constraints governing the ISP services domain have to be transformed into logical expressions that can then be expressed using the expressive syntax of OWL.

Although the work on the classification of ISP services and their specification using the service description template (as described in section 4 above) has been almost completed and fully documented, the work on the documentation of the exact rules that each service will have to comply with has not started yet. Since it is expected that this task will involve a great amount of work which will be continued even after the end of the project, a workaround has been used to ensure that OWL is adequate to express all possible rules and restrictions in the ISP world. A demo service was described and the following possible types of restrictions that the modeller could be asked to represent were identified:

1. The set of selectable enumerated values of a characteristic depends on the value of another characteristic.

2. The existence of a characteristic depends on the value of another characteristic.

3. The range of values of a characteristic depends on the values of another characteristic (very similar to case 1 but more complex to represent because deals with ranges).

4. Characteristics are mutually exclusive, but at least one must be present although they are indicated as optional.

The first type of restriction essentially constitutes an “if” statement: if the value of characteristic A is greater than 5, characteristic B can only take values 1, 2 and 3. The representation of an “if” statement can be easily achieved with the help of the logical operators AND, OR and NOT. If X then Y is equivalent to NOT X OR Y. Of course, OWL allows the definition of such a logical restriction.

As the same analysis has been performed for all the above types of restrictions, it is clear that OWL and the definition of the SKB ontology model satisfy
the requirements of the ISP domain as well as of the VISP use case called “validation of a service set”.

A technical restriction imposed by the current OWL specification is the lack of support for numeric ranges. While OWL has cardinality restrictions, datatype values cannot be further restricted. One of the solutions evaluated was to extend OWL files with URI references into an (external) XML Schema file. This means that a separate XML Schema file together with the OWL file will have to be maintained. Another solution is to define an extension ontology to be imported by the SKB ontology, which defines RDF properties that can be used to represent XML Schema facets. Other solutions are also being evaluated and hopefully the next OWL specification will be enriched with the capability of expressing restricted datatypes.

5.3 Use of the SKB Ontology

The use of the ontology comprises two concepts: how to manipulate the ontology and how to exploit it to enable the validation of a service set.

When manipulating the SKB ontology, Protégé allows for a series of actions regarding a definition of an ontology, such as to:
- create the ontology definition
- browse the ontology
- edit/update the ontology definition
- query the ontology
- store the ontology persistently
- apply various ontology checks
- compute the inferred taxonomy

The last two actions are performed with the assistance of an OWL reasoner, which can be combined with Protégé if it supports the DIG interface. An OWL reasoner provides the following standard inference services:
- Consistency checking: determines whether the ontology contains any contradictory facts.
- Concept satisfiability: determines whether it is possible for all classes to have any instances.
- Classification: computes the inferred class hierarchy.

In general, Protégé supports all features expected of an OWL editor. The SKB ontology will be manipulated through the use of Protégé, at least until the end of release 3 of the VISP software infrastructure. In the future it may be desirable to be able to perform certain actions on the ontology from the VISP platform itself.

The primary reason for the definition of the SKB ontology was to enable the implementation of the validation of a service set. When a customer requests one or more services from VISP, a sales representative creates a service set that comprises various services offered by partners in the cluster. After a service set has been created, it needs to be validated to ensure it can be instantiated. The sales representative uses the VISP GUI to click on the “validate” button to validate this service set.

The use of the SKB ontology allows for a more sophisticated validation strategy than the rather simple and incomplete validation undertaken in the first two releases of the VISP infrastructure software. The ontology is being used to enhance the first implementation, e.g. checking intra-service constraints between characteristics and values, checking inter-service constraints between services, characteristics and values, etc.

Compared with the current implementation, the additional necessary architectural element will be a framework enabling the manipulation and the usage of the ontology. Such a framework can be Jena (http://jena.sourceforge.net/index.html). Jena is a Java framework for building applications related to the Semantic Web in general. It includes:
- RDF API
- OWL API
- Rule-based inference engine
- In-memory and persistent storage
- SPARQL query engine (www.w3.org/TR/rdf-sparql-query/)

Its most useful feature for VISP is that it allows an ontology or a data set to be represented with an object model, as well as incorporating reasoning capabilities.

A main prerequisite before the validation can start is the existence of a consistent SKB ontology. A service description is added to the SKB ontology when it is approved in the VISP cluster with the appropriate restrictions. The next steps are then followed inside the implementation code of the validation of a service set:
1. The system reads the service set that is to be validated either through the repository containing information about the service sets and their component services or through an XML file.
2. It creates a model representing the data set of this service set.
3. It loads the ontology model of the SKB ontology. This can be read and created each time from the respective OWL file or it can be stored persistently inside a database.
4. It validates the data model against the existing ontology model, i.e. it performs a global check across the schema and instance data looking for inconsistencies.
5. The result will report whether the validation check has passed and detail any detected inconsistencies. This approach is quite simple and can be seamlessly integrated with the current implementation of the service set validation since Jena is a Java framework that can be used by the existing classes.

6 CONCLUSIONS

The VISP project has developed both a methodology for categorizing and describing ISP services and an ontology for the ISP service domain. The need to add semantic content to ISP service descriptions is fundamental for ISPs carrying out eBusiness for innovative, complex and composite ISP services.

The VISP ontology has been produced for a specific context and is being applied in this context. This is indeed a recommended approach when the aim is a lightweight ontology that can be cost effective to design, build and maintain (Alani et al., 2008). Ontologies do not need to be large and complex and the intention of the VISP project was innovative in that it has designed an ontology that could be part of the VISP platform in a running environment. Decisions taken in developing the ontology tended therefore to be pragmatic with the context, stakeholders and potential users in mind.

The intention is to further develop the results of the VISP project for commercial use in a virtual cluster of small ISPs. The ontology is an essential part of this aim and has been designed for such use. The work depicted here has built on existing research work in the area and has applied it to a specific context. This has required a lightweight approach avoiding any undue complexity and the development of a small and highly focussed ontology that fits in with the specific requirements of the VISP cluster and its partners. First results have shown that the ontology has met these requirements and that ISP composite services can be automatically validated with the ontology during service provisioning.

ACKNOWLEDGEMENTS

This work was carried out with partial funding from the EU through the IST project VISP (IST-FP6-027178). Further information on VISP can be obtained from www.visp-project.org.

REFERENCES


ON-DEMAND MOBILE CRM APPLICATIONS FOR SOCIAL MARKETING

Business and Technology Perspective

Viktor Kaufman
SAP Research, Karlsruhe, Germany
viktor.kaufman@sap.com

Yuri Natcheto
SAP Research, Montreal, Canada
yuri.natcheto@sap.com

Vasily Ponomarev
NPP Rumb, Moscow, Russia
vasily.ponomarev@gmail.com

Keywords: CRM, mobile, social marketing, collaboration, customer loyalty, mobile application, application design.

Abstract: Solutions for Customer Relationship Management (CRM) tend to evolve from the traditional passive recording of transactions between the Company and the Customer to agile optimized strategies for interaction and cooperation with the Customer. Despite considerable industry and research interest to the new "social" CRM 2.0, the gap between the vision and the reality is still quite large. Use of handheld devices deserves special attention in this respect. Although it can boost sales, customer satisfaction and customer loyalty, the amount of research work and the number of mobile applications in this area is low. In this paper, we focus on design and implementation of mobile social CRM solutions, which make extensive use of collaboration technologies. By means of simple business use-case scenarios, we argue that good understanding of business objectives, business processes, and technology issues, together, is crucial for development of compelling social CRM applications.

1 MOBILE SOCIAL CRM

Nearly every business needs some kind of CRM solution, even if it is not a specialized CRM software system. Banks, Insurance Companies, e-Businesses, non-profit and Public Sector organizations - all focus on serving their customers.

A large CRM component is usually integral part of enterprise business-software landscapes. SAP, Oracle, Salesforce.com, and others are the established proprietary software vendors. Smaller companies also process information that is typical for CRM, such as leads, contacts, accounts, promotions, sales orders. In any case basic data management, activities tracking, and reporting functionality implies high maintenance costs. Especially costly is customer-acquisition support.

As the efficiency of traditional mass-marketing, such as TV, radio and direct mail, constantly decreases, companies face the need to review their CRM strategy. One promising cost-effective and otherwise strategically interesting approach is called Social CRM or CRM 2.0 (Band, 2008). Social CRM recognizes the fact that mass advertising is not appealing enough to customers and advocates more attentive and interactive approaches to foster customer relations. Due to the rise of the Internet and increased level of public knowledge, customers will rather actively search for information they need. They will leverage their social networks and trust to their friends and other customers like them. When making a choice to buy a product or service, they will ask opinions of members of their social network (Webber, 2007). Social CRM extends the collaborative CRM and focuses more on support of
co-operation and customer-involvement aspects of customer relations.

With regard to CRM objectives, mobility brings a number of advantages. Due to the ubiquitous nature of the mobile communication (Virki, 2007), it is especially suitable for social marketing. Most often, consumers and sales agents wish to make decisions when they are away from their computers and have limited access to information directories. Leveraging broad availability of mobile devices can facilitate ease of relevant information access, advanced customer service, and field workers support (Sudan et al., 2007). In Section 4 we discuss how this eventually leads to CRM-effectiveness gains.

Mobile social marketing is a new concept, but its role rapidly becomes perceived as significant for CRM promotion and delivery strategies (Webber, 2007). We consider mobile social marketing a “must have” element of the future integrated CRM solutions. Possible concrete applications include managing social links on the mobile phone, receiving recommendations, consulting community opinions about products, tagging, rating, promoting, and ranking goods and services on the move. Retailers could get additional means to gather information about consumer preferences and forecast consumer needs. Advanced applications can take care of mobile access to a marketplace, where retailers and consumers search for products and services with the best quality-price ratio, place their orders directly in the back-end CRM system etc. The vast popularity of content-generation modes in the context of social networking web sites (Vaske, 2008) suggests support for mobile exchange of subjective information related to products and services. Software vendors already support some of these ideas, although the major business-acceptance breakthrough requires more time.

1.1 Market Overview

Jim Balsilie, co-CEO of RIM, Canadian wireless device company, says (Lomas, 2008) that the need to seamlessly integrate Web services and desktop applications onto handsets “is not a concept. This is a reality”. He further mentions: "Once social networking becomes a B2B phenomenon – not unlike IM and texting – I believe every single social-networking user will want a data plan”.

SAP and some other major vendors of CRM solutions closely follow the new developments. Recently, social-networking B2B approach has been chosen to drive adoption of SAP software by small and medium-sized businesses (Darrow, 2008). The influencers can get incentives for registering and closing leads in the company’s CRM system.

Recently (Oracle, 2008), Oracle announced Mobile Sales Assistant for company’s CRM on Demand solution that “changes the face of mobile CRM”. It features user collaboration with colleagues and customers, push-based architecture for BlackBerry®, one-click-away account information and customer contact information, and more for $30 per user per month.

During the iPhone™ Software Development Kit announcement, March 6th 2008, Salesforce.com demonstrated its ability to bring its innovative CRM on-demand services onto the iPhone platform and to provide new level user experience. Social mobile CRM applications are to be expected soon.

Microsoft has only recently entered into the CRM market with Microsoft Dynamics CRM 3.0. Their Mobile CRM solution works only on the Windows Mobile Pocket PC devices. For now, it provides salespeople with up-to-date information about their accounts and contacts, and helps them manage sales opportunities and track sales-related activities.

Arvato Mobile offers a set of advanced building blocks for mobile CRM. They provide for engaging user experience including games, send-a-friend, sponsored pop ups, communication via SMS, MMS and more. The tool set both aims at “customer-club” for customers and supports statistical data aggregation for CRM.

Kintera Inc. recently launched Kintera Sphere™ v8.0, company’s new social CRM system for non-profit organizations, providing a total view of the constituent’s relationship with the organization (CRM Today, 2008). The company claims to have considered factors such as enhanced trust, sense of belonging, instant gratification, emotional release, and sense of social impact. These features are definitely interesting for mobile applications too.

In the light of described mobile social CRM developments, the question remains – when and how will new solutions actually enable sales professionals to better understand and address their customers’ needs, anytime and anywhere. As companies move from products to solutions, the technology imperative is to enhance the business acumen and insight of the front line, which has not been the traditional goal of CRM. The business condition precedents to business-acceptance of innovative CRM solutions have to be understood in the first place see Sections 2 and 4. The challenges for "technology enablement” leading to effective
selling supported by CRM will be presented in Section 3. Our view is based on own applied research in the mobile CRM area described in Section 4 and summarized in Section 5.

2 YET ANOTHER CRM WITH MOBILITY SUPPORT?

Mobile access to back-end CRM systems is not new (Sadeh, 2002). In what respect are mobile social applications different? In our opinion, the answer can be found between the lines of the previous Section – it is the challenge to better meet business objectives by means of supporting and influencing complex human interactions. The idea is that customer and partners feedback, or ability to collaborate, and not so much stiff internal processes, drive customer relations, marketing and sales. Dynamically changing customer relations and the need to often review company’s self-perception in the market add to the overall complexity.

There is no one good solution, but poorly serviceable solutions are easily made. As the businesses remain sceptical, solution designers look for ways to control the complexity. In a nutshell, two approaches are under consideration by researchers and practitioners. The first becomes increasingly accepted and advocates use of general-purpose collaboration-support tools like Email, push-alerts, Instant Messaging, chat, mashups etc. alongside the company’s CRM processes. While there is nothing wrong with this approach, care is needed to decide whether it is appropriate. The costs of putting a solution into operation and maintenance costs can easily exceed the expected benefits (Galdy, 2008).

The second approach is to adapt and extend existing business processes. In (Band, 2007), the authors suggest the following four steps to build compelling CRM applications: 1) define and quantify business goals; 2) formulate CRM strategies and tactics; 3) establish appropriate CRM measures; 4) link CRM goals, strategies, and metrics. Forrester advocates another four-step strategy called POST (Bernoff, 2007): P) review the Social Technographics Profile (decide what is possible, customers are divided into groups like Creators and Inactives); O) pick an objective; S) choose a strategy; T) select and deploy appropriate technologies and measure results.

We argue that both approaches have advantages for mobile social CRM, but undervalue the intrinsically unstructured nature of customer interactions. The first approach needs dedicated control mechanisms to canalize and manage customer relations. The second approach contradicts with the idea of lightweight intuitive support for customer and partners interactions. Due to the formalization requirement, the processes can easily become incomprehensible or inflexible. We have chosen a mixed approach, which avoids undue quantification and specification of desired interactions, but puts strong emphasis on thought-out, informal, early and iteratively adaptable specification of the desired mobile social CRM solution.

We recognize that most important aspects of the social CRM are: 1) efficient cooperation with customers and partners; 2) efficient collaboration between CRM and non-CRM employees within the company; 3) supporting and enlarging existing relationships among customers and 4) shifting focus from the sales volume to better customer experience, compare also (Paterson, 2005). In fact, most mobile CRM solutions fail to address some or all of these issues. In our opinion, this is because of the mentioned complexity and practical difficulty of involving business analysts, technical experts, scientists, field workers, and customers early enough in the solution design.

We believe that transparent and flexible design principles constitute the best way to gradually accumulate expertise of the many people contributing to a compelling CRM solution. We follow an approach, where the following three steps are being largely addressed in parallel: 1) define and understand business objectives (market research, defining specific CRM measures, customer group focus etc. can be part of this step, but general understanding of implications of the business objectives is more important); 2) choose business processes (in the first place, we identify use-case scenarios to focus on); 3) address technology issues. This step is indispensable especially for mobile solutions. The architectural framework described in the next Section facilitates reuse of components, but some redesign and customisation are still necessary, depending on the desired business processes. Next, we proceed with step 3) and then exemplify steps 1) and 2).
3 TECHNOLOGY PERSPECTIVE: APPLICATION DESIGN

When designing mobile extensions to the Enterprise CRM applications, one must be careful to distinguish traditional software applications and specific functionality in a mobile-enterprise software context. It is impossible to deploy the entire business-software system onto a mobile device. The challenge is to transmit only the relevant business information and implement software functionality required for selected processes.

The known limitations of mobile devices are: connectivity, processing power, usability, security concerns, and memory requirements, but mobile devices also have some advantages, such as multi-modal input and high level of availability. All these considerations have to be taken into account. For social mobile CRM applications, there are additional requirements due to on-demand support of often changing customer interaction patterns.

With our business application framework (Natchetoi et al., 2008), we have implemented Mobile CRM client as a Java midlet and native Objective-C application for iPhone. In our solution, neither the business logic nor the user interface forms are hard-coded in the client application. Instead, the client application partially implements interpreters of open industry standards like SOAP, RDF, OWL and XForms. The application logic and user interface can be easily modified or augmented at low cost, since we are using standard formats and collaboration concepts such as mashups from the Web 2.0 tool set, see (Natchetoi et al., 2007) for details on our mobile business-oriented browser. For an example, user experience can be enhanced by adding XForms accessing back-end Business Objects exposed through SOA Web Services.

To lighten the application, only subsets of the mentioned standards are used. For example, a very limited version of the full BPEL interpreter is implemented on the mobile device, which enables basic composition of the workflow scenarios using local and remote services (Hirsh et al., 2006).

The components overview of our architecture is presented in Figure 1. In our Framework, the CRM business-data objects are being serialized, compressed and transmitted to the client side in the form of a compressed RDF messages. The information is stored in the local Persistent Data Store, also in the compressed RDF format, making it possible to store a significantly larger number of business objects as compared to a traditional file system or relational database. The Framework enables Web service calls and pro-active download of data required later on in the asynchronous business process. The client application uses the locally stored data to support off-line work.

Efficient connection to and data synchronisation with the back-end Enterprise system is very important for field workers. We use asynchronous, message-based communications for this, as they are a better fit in the mobile environment. The Smart Asynchronous Connection Manager (SACM) is a unified manager of asynchronous communications between and with mobile clients. The standard communication protocols supported by now are TCP/IP and SMS. We designed suitable API to asynchronously send messages from the client to the server, from the server to the client, and between the clients. The SACM does not explicitly distinguish clients and servers so that it also suitable for mobile Peer to Peer scenarios.

Our Mobile CRM prototype based on the introduced Framework enables sales agents to establish interactions using comprehensive persistent knowledge about the constituents and provides a centralized way to persist and manage customer
relationships information. Mobile CRM workers can update business information, browse reports and invoke Web Services remotely. They can access the CRM application functionality both online and offline. The mobile CRM has been developed for iPhone (Finkle, 2007) and for J2ME-enabled phones using the same Framework.

Figure 3: Sample screenshots of the mobile CRM application.

The Framework enables simplified adaptation or extension of existing applications. Beyond the basic Contacts, Leads, Opportunities, and Sales Orders management, we implemented some collaboration Services supporting multi-channel communication of sales agents, in a way inspired by ECOSPACE project (http://www.ip-ecospace.org). For now, we support Email and Fax access, and Instant Messaging integrated into CRM-processes context. In the next Section we describe further functionality relevant to mobile social CRM, which we could support in the future.

4 BUSINESS PERSPECTIVE: ENABLEMENT OF SALES PEOPLE AND CUSTOMERS

Social CRM is essentially about interaction with customers. The sales people and the customers are the prominent interacting parties. It is therefore natural to support them in the first place. This simple assumption already sets priorities for the functionality and the needed technology. We herewith assume that interactions drive the development and customization of CRM applications.

There are still different kinds of customers, sales people and interactions. Traditional CRM solutions would establish order by suggesting "best practice" processes out of the box. This is not enough, though, in the world of ever more rapidly changing requirements and increasing customer expectations. As described in Section 2, we have to focus on specific business objectives, using selected business processes.

The corresponding decisions include understanding of company's actual or desired core business strengths and deriving strategic objectives for interaction with customers. The objectives have to be detailed by taking into account customer profiles, relevant market researches, anticipated trends etc. They can be focused on a particular customer segment, technology, measurable performance indicators etc. In any case, the decisions should lead to a carefully selected set of desired customer-interaction-support scenarios, which have to be continuously validated against the business objectives and provide for differentiation from the competitors.

To test and further develop our mobile CRM application and our mobile Framework, described in Section 2, we were looking for some innovative scenarios, which we could implement in the near future. It became clear that we needed to define a compelling set of objectives first. Let us consider an example.

Imagine a Service provider company. It has to face harsh competition. To survive the next drop in prices, the company decides to save on traditional marketing and to leverage the knowledge of loyal customers about the company's quality services. The management wants customer retention to be addressed more; hopes to be able to provide added value especially for new customers; and strives to differentiate the company from the competition. To achieve their plans, the management considers the following use-case scenarios.

4.1 Set of Scenarios 1

The company needs ways to communicate to its loyal customers. Those customers are presumably rather busy and sceptical about advertisements. Therefore, for most scenarios described below, traditional communication channels through mail, email, and telephone would be considered annoying and would be too inefficient. The company decides to go for frank short conversations with rather obvious added value for the customer, and sees handheld devices to be best suited for this.

One option would be to cooperate with a mobile carrier and provide mobile services, including the ability to send SMS or other kinds of messages to the customers. Another option would be to provide
mobile applications for download. Also it is possible to cooperate with a handheld devices vendor. The last two options are much more flexible, compare below, and the corresponding applications would make use of the technical architecture described in the previous Section.

4.2 Set of Scenarios 2

The company is interested in most active, experienced, or otherwise influential persons among the customers. Determining such customers is to some extent similar to representing the customer base as a “social network”. The company can make an educated guess that loyal customers or customers with high amount of orders are potentially valuable collaborators. To facilitate their retention and to test the guess, it offers the selected customers to forward a small gift to a person of their choice (but not themselves). The gift can be limited Internet domain contract, free SMS pack, credit points for some services, etc. The choice of gifts can be made more intelligent later, as the “social network” is growing. The customers have the choice not to store the delivery address at the company, since the mere act of forwarding is enough for the company's purposes.

As soon as the influential customers are known, one can offer them certain incentives and ask to recommend the services. The company offers to conduct a counselling interview. The influential customer suggests time and enters contact details of her friend. The company can reassure the influencer of its quality counselling, including comparisons with similar services offered by competitors.

4.3 Set of Scenarios 3

The company might need some user feedback on provided services. One idea would be to offer a downloadable game, which becomes activated after the customer has completed a short satisfaction survey. Playing the game, the customer may have the option to share her view on possible improvements for company's services. For specific services, it may be possible to review other users' feedback, and even to discuss them in a kind of forum, adapted for mobile use. Considering the technical reality of mobile users is facilitated by the appropriate technical design.

Especially new customers would benefit from sharing their problems-SMS in the forum. This service can be supplemented by an Internet based online community.

4.4 Set of Scenarios 4

Selected “premium” users can earn gifts, reputation, or even money by helping out in the scope of their experience. The company provides them with a mobile tool, which effectively counts them to the product support team. They receive and answer questions, concerns and ideas of other users, answer and forward them. They can exclusively send their own requests to the company's support team members. The application architecture described before would even allow certain direct changes in the back-end systems on their behalf. Multimodal input would facilitate simple upload of a picture, video, or voice explanation to the designated online space, possibly, after review of company's responsible employees. The interested other customers can be notified directly of this event. Next, the chat functionality facilitates communication. To achieve even more customer loyalty, additional services of social value like exchange of contact details or introduction of customers with similar profile to one another can be offered.

4.5 Set of Scenarios 5

Sales agents use back-end services directly on their mobile phones. They also benefit from the emerging social space. They can tap the experience of “premium” customers or ask them for references. The sales agents can exchange thoughts and business content through chat and other functionality, which is deployed on their mobile phones. As a side effect, the company can perpetuate discipline in using back-end CRM system and, through collaborative style of work it can uphold the sense of mission for sales agents. Similarly to customer feedback, sales agents can communicate their concerns directly to the sales-support team.

4.6 Advanced Scenarios

To strengthen its market position by means of key differentiators, the company plans to provide special kinds of services, based on its customer-base information. The company's customer profiles include information on their mutual relationships and joint activities. It is now possible to derive or extract information, which can be used for strategic decisions. The customers can be divided into segments; the adoption level of selected scenarios can be measured etc. Some of the corresponding analysis results should be available on the mobile
phones for sales agents. Our framework described before allows to rather easily extending applications to include reporting functionality.

Given the metadata about joined CRM activities, mathematical analysis of gathered data can be performed. For example, business classifications can be tested and improved by means of probabilistic methods like dynamic Bayesian networks (DBN) and dynamic conditional random fields (CRF) (Lafferty et al., 2001). Promotion scenarios and playing with what-if scenarios becomes possible.

The “social network” data can be enriched with data shared on the social sites like MySpace or Facebook. Then, also cross-selling opportunities and more broad marketing can be pursued. Mathematical analysis can identify potential new customers.

4.7 Implications of Scenarios Use

The general message of this Section is to focus on customer and sales-agent needs as much as possible, and at the same time to achieve the company's objectives. Through more focused approach, flooding customers with irrelevant advertisements becomes superfluous. The described scenarios show how customer retention, market penetration, decision support, and powerful reporting can be achieved. Establishing relevant social processes would push the limitations of the traditional CRM applications far into the area of more educated Social CRM, with Mobility providing added value.

4.8 Experimental Evaluation

We are in the evaluation phase of the approach, presented in the paper. It is difficult to evaluate such characteristics as user’s trust quantitatively. We have tested our Mobile CRM client prototype with users from 11 different industries. Most of them have ranked the Social Marketing and Collaboration elements of the prototype as “somehow useful” and “very useful”.

In order to verify our ideas on Social Web-based Marketing approach we have also implemented a Web portal targeting the young people, students and young families living in Moscow. The portal, funded by the Moscow Government with highly dynamic content provides various information services to the citizens. The advantage of taking Social approach for building this portal is obvious because it is targeting different categories of users, living in the same city and often linked to each other.

The portal informs citizens about municipal programs as well as helps them to get education, find job, plan entertainment and receive consulting support from the social aid experts. The portal is a complex proprietary CRM system, treating citizens as clients, served by the municipal Government. It includes multiple elements of collaboration, smart content classification and auto-actualization. The content addressed to the different categories of the users passes through the actualization engine that is based on linguistics structure-based classification approach to direct content to the appropriate group of the users. (Ponomarev 2004) The feedback, collected from the users is processed in order to improve auto-actualization procedure in iterative way.

The results of three-year long experiments of using Social approach for this portal have been collected in order to evaluate the efficiency of the taken approach. However, the efficiency of such solution is difficult to express in numbers, because the goal of such Social portal is not the sales volumes but rather the trust and satisfaction of the users. The results of the customer satisfaction surveys, taken by the users during the period of two years indicate a strong positive shift in the “brand” loyalty and customer satisfaction. (Ponomarev 2005)

However, we are still looking into the different approaches to quantify user’s satisfaction and trust. We are also looking forward to implement mobile access to the Social portal for youth and make this feature available to all users.

5 SUMMARY

Mobile social CRM is a new concept that promises more focused, rewarding, engaging and powerful way to build long-lasting customer relationships. Use of Web 2.0 collaboration technologies can provide major benefits. These benefits come at a cost of more responsible application design. The CRM applications have to become agile with regards to the needed functionality. Mobility support is especially technically demanding.

Furthermore, regardless of the available technical expertise, the usefulness of applications has to be derived from thought-out business requirements. A formalisation of suitable business requirements for less structured collaboration and communication processes is difficult by definition. To circumvent this issue, we suggest following loosely defined steps suitable for unstructured nature of pervasive interactions in social CRM. In short, these steps define business objectives, business processes and enabling technology.
Grounded decisions on the overall business objectives, in the first place lead to requirements for a draft technical architecture with flexibility for the unstructured organic growth processes in mind. One has to consider expertise from relevant vertical divisions of the company, identify use-case scenarios in as many details as possible, consult potential users, and particularly take care of community support, make decisions on the scope of the scenarios to support, adjust the technical design to ensure adaptability with respect to the scenarios and provide the software lifecycle support. Many steps can be taken in parallel, but they all require continuous exchange between the stakeholders, and are best elaborated gradually. Transparent simple design principles constitute the best way to gradually accumulate expertise of the many people contributing to a compelling CRM solution.

In this paper, we exemplified the mentioned points. We considered scenarios to support proliferation of business for a Service provider. Due to flexible technical design, our prototype mobile CRM application can be extended to support many of the business scenarios at a very low cost. We outlined the important technical decisions.

We also discussed provision of advanced back-end functionality, such as analysing community structure and producing powerful reports. Such advanced functionalities can serve as a key differentiator from the competition.

REFERENCES

Band, W., 2008. The CRM 2.0 Imperative: Look To New Solutions To Keep Pace With Emerging Social Customer, Forrester Research, Inc.


Ponomarev V.V. 2004, Conceptual Model of the linguistic software complex of the expert search system with elements of the socio-psycho linguistic determination. Dialog MEPhI, p. 176

Ponomarev V.V.2005, Implementation of the new approach to automated reporting from the terminal stations of the municipal information system “Youth”: New software technologies, Moscow, MGIEM, p. 44-51.
E-COMPLEMENTARITY
The Link to e-Business Value

Pedro Soto-Acosta and Angel L. Meroño-Cerdan
Department of Management & Finance, University of Murcia, Campus de Espinardo, Murcia, Spain
psoto@um.es, angelmer@um.es

Keywords: e-Business, Resource-based theory, Internet, business value, Information technology.

Abstract: In recent years, much debate about the value of e-Business and information technology (IT) has been raised. Although the macro-level effect of IT and e-Business is undisputed, a question remains on whether e-Business can provide differential benefits to individual firms. In this sense, there is a need to further investigate whether and how e-Business creates value. To respond to this challenge, this paper develops a conceptual model, grounded in the resource-based theory, which analyzes the complementarity of Internet resources and e-Business capabilities as source of business value. This model posits three relationships: Internet resources and business value, internal e-Business capabilities and business value, and the complementarity of Internet resources and internal e-Business capabilities. To test hypotheses, a sample comprising 1,010 Spanish firms is employed. The results show that, as hypothesized, Internet resources per se are not positively related to business value and that internal e-Business capabilities have a positive significant impact on business value. In addition, the results offer support for the complementarity of Internet resources and internal e-Business capabilities as source of business value.

1 INTRODUCTION

The relationship between information technology (IT) and business value has been the subject of much research over the past decade. The results of these studies were varied and the term “productivity paradox” was coined to describe such findings. Nonetheless, recent studies have found positive and stronger linkages, and have attributed the productivity paradox to variation in methods and measures (Devaraj and Kohli, 2003)

Today IT is surpassing its traditional “back office” role and is evolving toward a “strategic” role with the potential not only to support chosen business strategies, but also to shape new business strategies (Henderson and Venkatraman, 1999). However, much debate about the value of IT and e-Business has been raised, due to the gap between e-Business investment and the lack of empirical evidence on e-Business value. Although showing recent signs of advance, much of the existing e-Business literature still relies, to a great extent, on case studies, anecdotes, and conceptual frameworks, with little empirical research directed to assessing the impact of IT on firm performance – especially in traditional companies (Brynjolfsson and Kahin, 2002). Case studies on firms such as eBay and Amazon show e-Business can create business value, but there is a question as to whether the lessons learned from these “Internet giants” are more widely applicable. At the same time, Carr’s assertions (2003), in his article “IT Doesn’t Matter”, have raised the discussion about the value of IT. Carr’s argument, in a few words, is that because every firm can purchase IT in the marketplace and because IT is now a commodity based on standards that all companies can freely use, IT is no longer a differentiating factor in organizational performance. What makes a resource truly strategic – what gives it the capacity to be the basis for a sustained competitive advantage - is not ubiquity but scarcity. Carr argues that no firm can use IT to achieve a competitive advantage over its competitors. Therefore, Carr concludes, firms should reduce spending on IT, follow rather than lead IT in their industry, and avoid deploying IT in new ways.

Most management information systems experts disagree with Carr’s assertions. However, his argument is appropriate when he points out that not all IT investments have strategic value. Some IT investments only allow firms to stay in business. The technology itself will rarely create superiority. For that reason, some research studies found that IT spending rarely correlates to superior financial
results (Hoffman, 2002). However, even though competitors may copy an IT innovation, relative advantage can be created and sustained where the technology leverages some other critical resource. Kettinger et al. (1994) draw a number of such complementary resources, such as size, structure, culture, and so on, that could make it difficult for competitors to copy the total effect of the technology. This complementarity of resources is a cornerstone of the resource-based theory and has been offered as an explanation of how IT has largely overcome its paradoxical nature and is contributing to business value (Bhatt and Grover, 2005; Clemons and Row, 1991).

Consequently, to respond to these challenges, this paper develops a conceptual model, grounded in the resource-based view (RBV) firms, to analyze the complementarity of Internet resources and capabilities as source of business value at the level of an individual firm. The analysis employs a large sample of companies from different industries for hypothesis testing. Moreover, although recent studies (Zhu, 2004; Zhu and Kraemer, 2005) have analyzed the relationship between e-Business capabilities and firm performance, very little work has been undertaken to identify Internet resources and capabilities. Similarly, the complementarity of Internet resources and capabilities has not been studied. The present study attempts to cover these gaps in the research.

The paper consists of six sections and is structured as follows: The next section reviews the relevant literature. In Section 3, hypotheses and research models are specified. Following that, the methodology used for sample selection and data collection is discussed. Then, data analysis and results are examined. Finally, the paper ends with a discussion of research findings, limitations and concluding remarks.

2 LITERATURE REVIEW

2.1 The RBV and e-Business

The RBV suggest that the effects of individual, firm-specific resources on performance can be significant (Mahoney and Pandian, 1992). The RBV generally tends to define resources broadly and include assets, infrastructure, skills, and so on. While resources serve as the basic units of analysis, firms create competitive advantage by assembling resources that work together to create organizational capabilities. Grant (1991) suggests that the capabilities of a firm are what it can do as a result of teams of resources working together. Teece et al. (1997) argued that capabilities cannot easily be bought; they must be built. Thus, building capabilities is not only a matter of combining resources; capabilities are rooted in processes and business routines. Also capabilities involve complex patterns of coordination between people and between people and other resources (Grant, 1991), and between an organization and other organizations. In this respect, Day (1994) describes capabilities as complex bundles of skills and accumulated knowledge, exercised through organizational processes, which enable firms to coordinate activities and make use of their assets. Day argues that capabilities and organizational processes are closely entwined, because capabilities enable the activities in a business process to be carried out. More recently, Makadok (2001) considers capability as a special type of resource. More specifically, he defines capability as an organizationally embedded non-transferable firm-specific resource whose purpose is to improve the productivity of the other resources possessed by the firm.

For the purposes of the present study, the above definitions of capability permit the identification of three important characteristics:

- Capabilities are rooted in processes and business routines, because it is capability that enables the activities in a business process to be carried out.
- Capabilities are firm-specific, while an ordinary resource is not. Because of this embeddedness, ownership of a capability cannot easily be transferred from one organization to another.
- The primary purpose of a capability is to enhance the productivity of the other resources that the firm possesses.

2.2 e-Business Resources and Capabilities

The RBV provides a solid foundation to differentiate between IT resources and IT capabilities and to study their separate influences on performance (Santhanam and Hartono, 2003). Based on this analysis, Bharadwaj (2000) suggested that if firms can combine IT related resources to create unique IT capabilities, they can improve their performance. IS researchers have followed this consideration of IT capability because competition may easily result in the duplication of investment in IT resources, and companies can purchase the same hardware and
software to remove competitive advantage (Santhanam and Hartono, 2003). In this respect, IS research offers a useful distinction between IT resources and IT capabilities. The former is asset-based, while the latter comprises a mixture of assets formed around the productive use of IT.

In general, IT resources are not difficult to imitate; physical technology is by itself typically imitable. However, firms may obtain competitive advantages from exploiting their physical technology in a better (and/or different) way than other firms, even though competing firms do not vary in terms of the physical technology they possess. IT resources are necessary, but not a sufficient condition, for competitive advantages (Clemons and Row, 1991). IT resources rarely contribute directly to competitive advantage. Instead, they form part of a complex chain of assets (IS capabilities) that may lead to better performance. Thus, some researchers have described this in terms of IT capabilities and argue that IT capabilities can create uniqueness and provide organizations a competitive advantage (Bhardwaj, 2000, Bhatt and Grover, 2005; Mata et al., 1995; Ross et al., 1996; Santhanam and Hartono, 2003).

Consequently, the present study seeks to demonstrate that although Internet resources (considered as physical IT) are not responsible for the creation business value, their complementarity with e-Business capabilities is critical to firm value.

2.3 Business Value from a Process Perspective

Although much research using the RBV has focused on an aggregated dependent variable, namely, firm performance, this may not be the best way to test the RBV (Ray et al., 2004). For example, because firms can have competitive advantage in some business activities and competitive disadvantage in others, examining the relationship between resources and capabilities associated with different processes within a firm and its overall performance can lead to misleading conclusions. Ray et al. (2004) proposed examining the effectiveness of business processes as a way to test the RBV logic. Another issue is that some IT investments may provide benefits after a certain period but increase operating costs in the short term. Thus, using firm performance at the macro level is meaningless and can again lead to misleading conclusions. These arguments lead to the conclusion that a process approach should be used to explain the generation of e-Business value within the RBV, and this is the approach adopted in the present study. The present research uses the effectiveness of online procurement to measure e-Business value. The business value of this process is discussed below.

E-Procurement, or buying online, can potentially provide distinct value propositions to the firm. These come from the reduction of procurement and inventory costs, as well as strategic networks with suppliers that allow effective and efficient supply chain management (SCM). With regard to procurement costs, Kaplan and Sawhney (2002) indicated that buying in e-marketplaces considerably reduces transaction costs. With regard to strategic links and SCM, Internet technologies can enhance SCM decision making by enabling the collection of real-time information, and access to and analysis of this data in order to facilitate collaboration between trading partners in a supply chain. In this sense, Frohlich and Westbrook (2002) showed the importance of linking customers and suppliers together in tightly integrated networks. As a result of e-Procurement, the collection of real-time information on demand is possible and, more importantly, products and services are delivered quickly and reliably when and where they are needed (Frohlich, 2002).

In sum, e-Business value may lead to improved performance on the part of the firm in procurement. Although it could be argued that customers, suppliers and/or the firm’s wider value network can benefit from online procurement, this study focuses on analyzing business value at the level of an individual firm.

3 DEVELOPMENT OF HYPOTHESES

This section develops hypotheses for the present study, drawing on the existing information systems and e-Business literature. Three relationships will be explored: Internet resources and business value, internal e-Business capabilities and business value, and the complementarity of Internet resources and internal e-Business capabilities (see Figure 1).
3.1 Internet Resources and Business Value

Firms obtain competitive advantages on the basis of corporate resources that are firm specific, valuable, rare, imperfectly imitable, and not strategically substitutable by other resources (Barney, 1991). IT resources are easy to duplicate, and, hence, IT resources per se do not provide competitive advantages (Santhanam and Hartono, 2003). Although IT infrastructure is argued to be valuable, it is not a source of competitive advantage (Bhatt y Grover, 2005). Thus, IT infrastructure will rarely lead to superior performance. Similarly, Internet resources – as defined above – are not difficult to imitate. In general, Internet technology is by itself imitable. If one firm can purchase certain Internet technologies and thereby implement some strategies, then other firms should also be able to purchase these technologies, and thus such tools should not be a source of competitive advantage. Furthermore, as the diffusion of the Internet continues, the ability of proprietary IT to be a source of competitive advantage continues to be eroded. These arguments suggest that Internet resources may not have a significant impact on business value. Thus, the following hypothesis is proposed:

Hypothesis 1: There is no relationship between Internet resources and business value

3.2 Internal e-Business Capabilities and Business Value

Investing in IT is not a necessary nor sufficient condition for improving firm performance, since IT investments might be misused (Tallon et al., 2000). In this sense, IT assets cannot improve organizational performance if they are not used appropriately. However, when used appropriately IT is expected to create intermediary effects, such as IT being embedded in products and services, streamlined business processes, and improved decisions, which can be expected to have an influence on the performance of the firm (Ravichandran and Lertwongsatien, 2005).

Grant (1991) and Makadok (1991) emphasize that while resources by themselves can serve as basic units of analysis, firms create competitive advantage by assembling these resources to create organizational capabilities. Makadok states that these firm-specific capabilities, embedded in organizational processes, provide economic returns because that firm is more effective than its rivals in deploying resources. IS researchers have adopted this capability logic of resources by arguing that competitors may easily duplicate investments in IT resources by purchasing the same hardware and software and, hence, IT resources per se do not provide competitive advantages. Rather, it is the manner in which firms leverage their IT investments to create unique capabilities that impact firm performance (Clemons and Row, 1991; Mata et al, 1995). Thus, it is expected that internal e-Business capabilities are positively associated with business value. The following hypothesis incorporates these expectations:

Hypothesis 2: There is a positive relationship between internal e-Business capabilities and business value

3.3 The Complementarity of Internet Resources and Internal e-Business Capabilities

Although there is research that posit a direct relationship between IS resources/capabilities and firm performance (Bharadwaj, 2000; Feeny and Willcocks, 1998; Santhanam and Hartono, 2003), others have questioned the direct-effect argument and emphasized that IS resources/capabilities are likely to affect firm performance only when they are deployed to create unique complementarities with other firm resources (Clemons and Row, 1991; Powell and Dent-Micalef, 1997).

Firm resources are considered complementary when the presence of one resource enhances the value or effect of another resource (Ravichandran y Lertwongsatien, 2005; Zhu, 2004). For example, the complementarity between online offerings and offline assets is the essence of “clicks-and-mortar” companies. Customers who buy products over the Internet value the possibility of getting support and service offered through bricks-and-mortar retail outlets, including the convenience of in-store pickup and return (Zhu, 2004). Hence the RBV highlights the role of complementarity as a source of value creation in e-Business, though is not the only source as suggested by Amit and Zott (2001). As mentioned earlier, Internet resources are not difficult to imitate and per se do not provide competitive advantages. However, having a proper Web infrastructure may facilitate the internal processing of online operations and this way influence positively firm performance. That is, the fact of possessing an adequate Web infrastructure can be critical for the influence of internal e-Business capabilities on business value. Thus, the following hypothesis is proposed:
Hypothesis 3: The complementarity between Internet resources and internal e-Business capabilities explains variations in business value.

4 METHODOLOGY

4.1 Data

The data source for the present study is the e-Business W@tch survey 2004, an initiative launched by the European Commission for monitoring the adoption of IT and e-Business activity. The decision-maker targeted by the survey was normally the person responsible for IT within the company, typically the IT manager. Alternatively, particularly in small enterprises without a separate IT unit, the managing director or owner was interviewed. The population considered in this study was the set of all enterprises which are active at the national territory of Spain and which have their primary business activity in one of ten sectors considered. The sample drawn was a random sample of companies from the respective sector population with the objective of fulfilling strata with respect to business size. A share of 10% of large companies (250+ employees), 30% of medium sized enterprises (50-249 employees) and 25% of small enterprises (10-49 employees) was intended. The number of firms totalled 1,010. 91.1% of firms were small and medium-sized enterprises (less than 250 employees) and each sector considered had a share of around 10% of the total sample.

With regard to respondents’ titles, 54.4% were IS managers, nearly 20% were managing directors, and 12.1% were owners. The dataset was examined for potential bias in terms of the respondents’ titles. Since respondents included both IT managers and non-IT managers, one could argue that IT managers may overestimate e-Business value. To test this possible bias, the sample was divided into two groups: IS managers (head of IT/DP and other IT senior managers) versus non-IS managers (owner, managing director, strategy development and others). One-way ANOVA was used to compare the means of factor scores between the two groups. No significant differences were found, suggesting that the role of the respondents did not cause any survey biases.

4.2 Measures of Variables

Measurement items were introduced on the basis of a careful literature review. Confirmatory factor analysis (CFA) was used to test the constructs. Based on the CFA assessment, the constructs were further refined and then fitted again. Constructs and associated indicators are listed in the Appendix and discussed below.

- Internet resources construct. This construct represents the adoption of physical Internet technologies. In this sense, respondents were required to assess the presence of four Internet tools: website, Intranet, Extranet and LAN (local area network).
- Internal e-Business capabilities. This construct represents the use of online technologies for supporting internal business processes.
- Business value. As discussed earlier in section 2.3, the present research uses the effectiveness of e-Procurement for measuring business value. That is, business value is assessed through the business impact of purchasing online.

4.3 Instrument Validation

CFA using AMOS 4.0 was conducted to assess empirically the constructs theorized. Multiple tests on construct validity and reliability were performed. Model fit was evaluated using the maximum likelihood (ML) method. The measurement properties are reported below.

- Construct reliability. All constructs had a composite reliability over the cut-off of 0.70 (Straub, 1989), and also the average variance extracted for all exceeded the preferred level of 0.5 (Churchill, 1979).
- Content and construct validity. Content validity was verified by checking the meanings of indicators and by a careful literature review. Construct validity is the extent to which a construct measures the concepts that it purports to measure (Straub, 1989). It has two components: convergent and discriminant validity. After dropping insignificant items, all estimated standard loadings were significant, suggesting good convergent validity. To assess the discriminant validity, all constructs met this criterion.

Table 1 lists several goodness-of-fit statistics to assess how well specified models explain the observed data. The insignificant p-value (p = 0.187) for the chi-square statistics implied good absolute fit. The root mean square error of approximation (RMSEA) is was below the cut-off value 0.08 suggested by Browne and Cudeck (1993). Five incremental fit indices were all above the preferred level of 0.9 (Gefen et al., 2000).
Table 1: Measurement Model Fit indices.

<table>
<thead>
<tr>
<th>Goodness-of-Fit Indices</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Chi-Square</td>
<td>66.054</td>
</tr>
<tr>
<td>p-value</td>
<td>0.246</td>
</tr>
<tr>
<td>RMSEA</td>
<td>0.032</td>
</tr>
<tr>
<td>Normed Fit Index (NFI)</td>
<td>0.971</td>
</tr>
<tr>
<td>Relative Fit Index (RFI)</td>
<td>0.955</td>
</tr>
<tr>
<td>Incremental Fit Index (IFI)</td>
<td>0.997</td>
</tr>
<tr>
<td>Tucker-Lewis Index (TLI)</td>
<td>0.995</td>
</tr>
<tr>
<td>Comparative Fit Index (CFI)</td>
<td>0.997</td>
</tr>
</tbody>
</table>

5 EMPIRICAL RESULTS

Hypotheses were tested using hierarchical regression analysis. Table 2 shows Internet resources construct is not statistically significant, whereas e-Business internal capabilities construct is positive and significant (regression 1), as predicted. Regression 2 includes both main and the interaction effect. The interaction effect between Internet resources and internal e-Business capabilities was found significant, thus, supporting the complementarity of Internet resources and internal e-Business capabilities. To further test the significance of the interaction effect, the incremental $R^2$ between the full model (with interaction term) and the partial model (without the interaction terms) was compared. The result is reported in the lower rows of table 7. In regression 2, the incremental $R^2$ was 0.026, meaning that approximately an additional 3 percent of explained variance has resulted from the inclusion of the interaction effect with respect to regression 1. To compare the partial model against the full models, a Wald test was performed and the differences were found to be statistically significant. Based on this, the partial model was rejected in favour of the full model (Greene, 2000).

Table 2: Complementarity results: Impact on procurement.

<table>
<thead>
<tr>
<th>Independent variables</th>
<th>Regr. 1</th>
<th>Regr. 2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Internet resources (IR)</td>
<td>0.083</td>
<td>0.099</td>
</tr>
<tr>
<td>Internal EB capabilities (IEBC)</td>
<td>0.196***</td>
<td>0.151**</td>
</tr>
<tr>
<td>IR * IEBC</td>
<td>-</td>
<td>0.167**</td>
</tr>
<tr>
<td>F</td>
<td>6.595***</td>
<td>6.633***</td>
</tr>
<tr>
<td>R2</td>
<td>0.047</td>
<td>0.069</td>
</tr>
<tr>
<td>$\Delta R^2$</td>
<td>0.026**</td>
<td></td>
</tr>
</tbody>
</table>

Through this analysis, hypotheses H1, H2 and H3 found support.

6 DISCUSSION

The results showed that Internet resources are not positively related to business value. This finding is not surprising, since competitors may easily duplicate investments in IT resources by purchasing the same hardware and software, and hence IT resources per se do not provide better performance (Santhanam and Hartono, 2003). This can be explained through the RBV, because IT is not considered a resource that is difficult to imitate; IT is by itself typically imitable. This result supports the findings of recent research (Batt and Grover, 2005) that did not find evidence of a positive link between IT quality and firm performance. Similarly, Powell and Dent-Micallef (1997) showed that IT by itself cannot be a source of competitive advantage. Thus, our results confirm that Internet technology by itself will rarely create business value.

Furthermore, results demonstrate that there is a positive relationship between internal e-Business capabilities and business value. Our findings confirm the existing empirical literature. Bharadwaj (2000) and Santhanam and Hartono (2003) found that firms with superior IT capability do indeed exhibit superior firm performance. Ravichandran and Lertwongsatien (2005) showed that an organization’s ability to use IT to support its core competences depends on IS capabilities. Thus, even though competing firms do not vary in terms of the IT they possess, IS capabilities are rooted in processes and business routines and provide competitive advantage. In this sense, the results of the present study support the proposition that internal e-Business capabilities are positively associated with business value.

Finally, the empirical results offer support for the complementarity of Internet resources and internal e-Business capabilities. The RBV highlights the role of complementarities between resources as a source of business value. Researchers such as Steinfield et al. (1999) suggest that e-Business value can come from synergies between online and offline presence. In this sense, using case studies, they showed the lack of exploitation of these synergies in SMEs. Zhu (2004) developed a study which evaluates the impact of e-commerce and IT on firm performance (financial measures), studying both the main effects and the interaction effect of e-commerce and IT on firm performance. Our results support the RBV and e-Business literature, therefore, it can be concluded that having an adequate Internet infrastructure can be critical for the impact of internal e-Business capabilities on business value.
7 CONCLUSIONS, LIMITATIONS AND FUTURE RESEARCH

In recent years, much debate about the value of IT and e-Business has been created, due to the gap between e-Business investment and the lack of empirical evidence on e-Business value. Thus, today IS researchers face pressure to answer the question of whether and how e-Business creates value. The complementarity of resources is a cornerstone of the resource-based theory and has been offered as an explanation of how IT has largely overcome its paradoxical nature and is contributing to business value (Bhatt and Grover, 2005; Clemons and Row, 1991). Thus, to respond to these challenges, this study developed a conceptual model, grounded in the resource-based view (RBV) firms, to analyze the complementarity of Internet resources and e-Business capabilities as source of business value at the level of an individual firm. The analysis employed a large sample of companies from different industries for hypothesis testing. Broadly, this research offers several contributions: (1) it identifies Internet resources and internal e-Business capabilities; (2) it shows that Internet technology by itself will rarely create business value; (3) it sheds light on the complementarity of Internet resources and internal e-Business capabilities as source of business value.

While the contributions of the present study are significant, it has some aspects which can be addressed in future research. First, the sample used was from Spain. It may be possible that the findings could be extrapolated to other countries, since economic and technological development in Spain is similar to other OECD Member countries. However, in future research, a sampling frame that combines firms from different countries could be used in order to provide a more international perspective on the subject. Second, the business value measure is subjective in the sense that it was based on Likert-scale responses provided by managers. Thus, it could also be interesting to include objective performance data for measuring business value. Third, the key informant method was used for data collection. This method, while having its advantages, also suffers from the limitation that the data reflects the opinions of one person. Future studies could consider research designs that allow data collection from multiple respondents within an organization. Fourth, this research takes a static, cross-sectional picture of capabilities, which makes it difficult to address the issue of how capabilities are created over years. A longitudinal study could enrich the findings.

REFERENCES


Journal of Management Information Systems, 16, 137-165.


APPENDIX: MEASURES

Internet Resources:
- Does your company have a website? (Y/N)
- Does your company use an Intranet? (Y/N)
- Does your company use an Extranet? (Y/N)
- Does your company use a LAN? (Y/N)

Internal e-Business capabilities:
- Do you use online technologies to share documents between colleagues or to perform collaborative work in an online environment? (Y/N)
- Do you use online technologies to track working ours and production time? (Y/N)
- Do you use online technologies to support human resources management? (Y/N)
- When an online order comes, is the order fully integrated with the back-end system? (Y/N)

Business value: e-Procurement effectiveness
- What effect has online procurement on the procurement costs? (1-5)
- What effect has online procurement on your relations to suppliers? (1-5)
- What effect has online procurement on the costs of logistics and inventory? (1-5)

Note. (Y/N), dummy variable; (1-5), five-point Likert-type scale.
A RESEARCH MODEL OF CUSTOMER RELATIONSHIP MANAGEMENT SYSTEMS FOR MOBILE DEVICES

Description of a Research Model about Customer Relationship Management Projects

Rebecca Bulander
Institute AIFB, University of Karlsruhe, Englerstr. 11, 76 128 Karlsruhe, Germany
Email: rebecca.bulander@aifb.uni-karlsruhe.de

Keywords: Customer Relationship Management, CRM, IT, CRM systems, CRM projects, mobile devices, mobile actors, business to business.

Abstract: In literature we can find evidence about a high number of failed CRM projects. However concerning CRM projects in combination with the implementation of mobile devices for field staff in sales and services there aren’t many results published. So there are missing information about the success of such projects, the success factors and the objectives. This paper presents a research model about such CRM projects. A CRM project in this paper is defined as an implementation of a CRM system together with mobile devices. The research model will give further insight into the success factors of such projects and the achieved objectives.

1 INTRODUCTION

Several studies (e. g. Alt 2004; Thomson et. al 2002; Ryals 2000) document the failure of CRM projects in the past. But concerning CRM projects related with the implementation of mobile devices for field staff in sales and services we find nearly no information so we can make no statement about how many of these projects failed. To gain more information about this and to prevent such CRM projects from failing it seems reasonable to inspect such projects. Therefore it is important to know what are the success factors and the achieved objectives of such projects. For this reason a research model will be given in this paper. The research model focuses on companies which sell their products and services to other companies – this is the so called business to business (B2B) sector. The B2B sector is more appropriate because in this sector we find a higher number of field staff working with mobile devices in combination with a CRM solutions than in the B2C sector.

For a better understanding there will be given some definitions. Customer Relationship Management (CRM) is a customer oriented strategy which is supported by information technology (IT) (Hipner 2004). CRM “[...] is the infrastructure that enables the delineation of and increase in customer value, and the correct means by which to motivate valuable customers to remain loyal – indeed, to buy again” (Dyché 2002, p. 4).

The information technology respective the CRM systems help to enhance the customer relation and to standardise the CRM activities and processes in fields of marketing, sales and service. This paper is focussing on CRM systems integrating actors using mobile devices. In terms of our research mobile devices are mobile handheld devices like cellular phones, personal digital assistants (PDA) and Smartphones but also on-board vehicle computers or notebooks.

A CRM project is a project which includes the implementation of a CRM system together with mobile devices but also the adoption of a customer oriented strategy. A project is a timely restricted endeavour to archive stated objectives; it has an assigned budget.

CRM systems together with mobile devices expand the channels of CRM to the customer (Sundararajan 2002) and assist both actors of CRM – employees of the suppliers and of the customers using mobile devices via public and private communication networks. Because the employees of the supplier (e. g. field manager gets mobile access to customer information) and/or of the customer (e. g. customer receives flight information on his cellular phone) can be in the working state mobile or
stationary we can differentiate between four CRM scenarios like shown in Figure 1 (Hampe 2002). In the mobile state the mobile actor is not located within its homelike infrastructure. The research model will enclose the three scenarios “classical CRM system”, “field staff” and “mobile link”. The scenario “mobile services” covers other aspects of CRM like mobile marketing and needs therefore other IT infrastructure.

<table>
<thead>
<tr>
<th>Working state of an employee</th>
<th>Customer</th>
</tr>
</thead>
<tbody>
<tr>
<td>stationary</td>
<td>stationary</td>
</tr>
<tr>
<td>mobile</td>
<td>mobile</td>
</tr>
<tr>
<td>stationary</td>
<td>Classical CRM system</td>
</tr>
<tr>
<td>mobile</td>
<td>Field staff</td>
</tr>
<tr>
<td>Supplier</td>
<td>Mobile Link</td>
</tr>
</tbody>
</table>

Figure 1: CRM scenarios using mobile devices.

2 LITERATURE REVIEW

Up to now, little research has been conducted in this described field of research (Salomann et al. 2005). To get a better overview about the research that has already been done, they will be grouped by the following aspects:

- Author and year
- Objective, research item and research focus business to business or business to customer sector
- Method of data analysis
- Database (sample size, kind of questioning, kind of investigation and countries in which the enterprises of the study are)
- Success factors of the CRM project
- Achieved objectives

The success factors are grouped by:

- CRM as strategy
- Implementation of CRM system
- Implementation of mobile information technology (mobile IT)

Also the achieved objectives of the CRM projects are sub-classified after the following aspects:

- related to the enterprise
- containing information about the objectives of the field staff

The last point is important because this information gives more insights about the result of an implementation of mobile devices. The last column in Table 1 shows if the research about the relationship between the success factors and the objectives is part of the study. This aspect is described with “connection between the variables”. All studies concerning this research topic are listed in Table 1.

As can be seen in Table 1 there is no study which covers all listed aspects. Most of the studies focus on the sector of B2C or B2B and B2C. There is only one study that focuses only on B2B. No study covers the aspect of mobile information technology and the achieved objectives in field staff. Some studies are influenced by a company which delivered the CRM systems or which did the consulting during the CRM projects.

According to the described research gaps there is a need of more research concerning mobile information technology in combination with CRM systems and CRM strategy. It is also necessary to analyse the objectives which are achieved by implementing CRM systems together with mobile devices by looking at the field staff. Therefore a conceptual development will be given and a research model will be presented.

3 CONCEPTUAL DEVELOPMENT

In the next paragraphs the research model in Figure 2 will be explained. The ellipses represent latent variables; these are the constructs of the model. The relationships between the variables are shown by the arrows between the ellipses. Each arrow is marked with ‘H’ and a number, which stands for the number of the hypothesis. The whole model contains 20 hypotheses. As shown in Figure 2 the success factors are on the left side and the objectives on the right side of the figure. According to Table 1 the success factors are divided into “CRM as strategy” and “implementation of CRM together with mobile IT”. The objectives are categorized into three groups: The first group is the overall satisfaction with the CRM project. The remaining groups are the objectives achieved in the whole enterprise. They are measured in four dimensions. The objectives referring to the field staff and the use of mobile devices are measured in three variables, also called dimensions.
Table 1: Selected studies on CRM systems.

<table>
<thead>
<tr>
<th>Author and year</th>
<th>objective, research item and research focus business to business (B2B) or business to customer (B2C)</th>
<th>method of data analysis</th>
<th>database</th>
<th>success factors</th>
<th>achieved objectives</th>
<th>Connections between variables</th>
</tr>
</thead>
<tbody>
<tr>
<td>Alt, Pushmann, Österle 2004</td>
<td>- Success factors of CRM projects - Research item: enterprises - B2C</td>
<td>Explorative analysis and benchmarking</td>
<td>- N = 6 - Personal interviews - Half standardized questionnaire - Countries: Germany, Switzerland, USA</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Greve 2006</td>
<td>- Success factors of the implementation of CRM systems - Research item: enterprises which are consulted by the company Accenture - B2C</td>
<td>Qualitative analysis, descriptive statistics, PLS analysis</td>
<td>- N = 90 - Written questioning - Standardized questionnaire - Countries: Europe, mostly Germany</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Jayachandran et al. 2004</td>
<td>- Influence of the CRM information process and the use of CRM technology to the customer relationship performance - Research item: enterprises - B2B and B2C</td>
<td>Qualitative analysis, factor analysis</td>
<td>- N = 172 - Written and online questioning - Standardized questionnaire - Country: no information</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reinartz, Krafft, Hoyer 2003</td>
<td>- Success factors and objectives of CRM implementations - Research item: enterprises - B2C</td>
<td>Qualitative analysis, descriptive statistics, PLS analysis</td>
<td>- N = 211 and N = 95 - Written and online questioning - Standardized questionnaire - Countries: Germany, Austria, Switzerland</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Selchert 2005</td>
<td>- Measurement of success of CRM projects in companies which have chosen CRM systems of SAG AG - Research item: enterprises - B2B and B2C</td>
<td>Descriptive statistics, Key performance indicators</td>
<td>- N = 32 - Telephone interviews - Standardized questionnaire - Countries: Germany, Austria, Switzerland</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Zahay, Griffin 2002</td>
<td>- Investigation about the customer based performance and the business growth - Research item: enterprises - B2B</td>
<td>Quantitative analysis, descriptive statistics, factor analysis</td>
<td>- N = 206 - Telephone interview - Standardized questionnaire - Countries: USA</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
3.1 Success Factor CRM Performance

The variable CRM performance stands for the ability of an enterprise to adapt the CRM strategy in its enterprise strategy and the CRM concepts and activities in its marketing, sales and service concepts. To measure the CRM performance three constructs are needed. Some aspects of these constructs are transferred from the study of Reinartz 2004. But the constructs of this study have to be converted from the B2C sector to special aspects of the B2B sector. Therefore the consciously determination of a customer relationship is not part of this model (Rapp 2000). So we get two antecedents for CRM performance.

First there are the actions to acquire new customers and second there are the actions to force the relationship to profitable customers and also to recommit inactive profitable customers (Reinartz 2004).

The actions to acquire new profitable customers and to build a relationship to the already profitable customers haven an influence on the CRM performance of an enterprise (hypotheses 1 and 2).

3.2 Success Factor Implementation of Information Technology

The variable implementation of information technology stands for the support of all customers related and oriented activities by IT. Especially in CRM are many activities, which wouldn’t be possible in an efficient way without IT – e. g. personalized advertising and newsletters. Therefore you find already in the definition of CRM the importance of IT (Hippner 2004).

To get an IT that covers all aspects of CRM there is a combination of the implementation of mobile IT and of a CRM system necessary. Therefore the variable implementation of information technology has two antecedents.

Only if the mobile technology and all other channels to the customer are integrated into the CRM system to ensure a bidirectional data flow on a consolidated customer data base all CRM processes and activities can be properly supported. This aspect is shown by hypotheses 3 and 4.
3.3 CRM Performance and Implementation of IT as Antecedents of Satisfaction

To measure enterprises satisfaction with the results of its determined CRM project in combination with the usage of mobile devices the variable “satisfaction with the CRM project” will be used. This variable shows the overall impression and covers the success of the CRM project over all. Therefore also the question is formulated, if the enterprise would repeat the CRM project.

The satisfaction with the CRM project is influenced by the two variables CRM performance and implementation of IT (see also hypotheses 5 and 6 in Figure 2).

3.4 CRM Performance and Implementation of IT as Antecedents of Achieved Objectives

Unlike the variable satisfaction with the CRM project the variables to measure the achieved objective cover certain aspects of CRM projects. Therefore a pre-study was undertaken to collect the objectives of CRM projects for mobile devices (Hartel 2006). The results of this pre-study have been grouped by topics. According to this pre-study we can differ between objectives related to the enterprise and to the field staff. Furthermore they can be assigned to sub-groups – so called dimensions.

The achieved objectives concerning the whole enterprise are grouped by four dimensions concerning the following aspects:

- finance (e.g. cost reduction or increase of up- and cross-selling)
- customer (e.g. improvement of customer communication or reduction of complaints)
- employees (e.g. improvement of the communication between employees)
- processes (e.g. improvement of effectiveness in sales)

The achieved objectives concerning the field staff using mobile devices are grouped by the following three dimensions:

- time and costs (e.g. reduction of cycle time or time saving by administrative work for sales people)
- quality (e.g. improvement of data and information quality for sales people)
- process for mobile devices (e.g. increase of process automation for sales and service people)

The antecedents to these seven constructs are the two variables “CRM performance” and the “implementation of IT”. Hence the relationship between these two variables and the variables in the certain dimensions are described in the hypotheses 7 until 20 (see Figure 2). All hypotheses in the research model show the following description – for example H7:

“The CRM performance has an influence on the achieved objectives of the enterprise in the dimension finance.”

4 RESEARCH METHOD AND FURTHER STEPS

4.1 Research Method

The proposed research model and its hypotheses represent a structural equation path model. This model can be tested by the Partial Least Squares (PLS) analysis. The PLS procedure was invented by Hermann Wold. It is a second-generation multivariate technique which has the ability to model latent constructs under conditions of non-normality (Chin 1999).

4.2 Further Steps

To prove the proposed research model further empirical research is needed. Therefore the variables have to be operationalized by indicator variables. After this the concept of the study has to be planned. This covers the aspects like: data base, sample size, kind of questioning etc. Each indicator variable refers to a question in the questionnaire.

Because of the different market of CRM systems only enterprises of one CRM market should be chosen for interviews – e.g. American CRM market or German speaking CRM market. Thus an influence by the market to the study results can be excluded.

By choosing the enterprises for interviews there should be avoided that there is any influence by a special CRM software product or a consulting company during the CRM project. The chosen enterprises should be free from such influences.

It is also important to select companies which have conducted a CRM project and completed it. Therefore it can be ensured that first influences as
well as long term influences of the CRM project for example in the dimension finance can be observed.

5 SUMMARY

The described research model shows a concept to analyze CRM projects in enterprises of the business to business sector. The model covers aspects of CRM strategy and CRM concepts as well as the support of the implemented information technology respective CRM systems. The model points out that the CRM performance and the IT have an influence on the achieved objectives of an CRM project. So the model shows that IT is not the only but an important factor to the success of such a CRM project.

The most important points of the model are that the variable “implementation of mobile information technology” covers all aspects for the use of mobile devices for field staff in sales and services. It is also important that the objectives of the field staff are measured and analyzed by three dimensions concerning aspects like time and costs, quality and processes for mobile devices.

REFERENCES

POSTERS
A STUDY OF INNOVATION DIFFUSION OF ELECTRONIC PATIENT RECORDS FOR SUPPORTING MEDICAL PRACTICE

Vincent Cho and Geoffrey Lieu

Department of Management and Marketing, The Hong Kong Polytechnic University, Hung Hom, Hong Kong

msvcho@polyu.edu.hk, mslieu@polyu.edu.hk

Keywords: Innovation diffusion, electronic patient record.

Abstract: This paper proposes a study on the underlying factors affecting the adoption, routinization and infusion of electronic patient record in the clinics of Hong Kong. We suggest using a focus group to identify the potential antecedents for the three stages of innovation diffusion (adoption, routinization and infusion). Then a theoretical framework based on the antecedents and their impact on innovation diffusion will be layout. It will be verified upon a survey sending to the medical practitioners in Hong Kong.

1 INTRODUCTION

We are well into the digital information age. Digital communications and information resources affect almost every aspect of our lives – business, finance, education, government and entertainment. Clinical practice is highly information intensive, but it is one of the few areas of our society where computer access to information has had only limited success.

Most IT practices in health care by physicians have been applied to office management in areas related to accounting of the business, and the scheduling on patients’ booking. The adoption of Electronic Patient Records (EPRs) - medical computerized systems that organize the information on a patient’s treatment, diagnosis and results from laboratory and other testing – appears limited. The possibility of instant, universal access to up-to-the-minute, accurate patient information is a goal that is actively sought throughout health services organizations. It is increasingly recognized that EPRs bring along the quality benefits of electronic documentation and viewing, prescription and test ordering, care management reminders, and messaging, among other medical systems. Thus, EPRs are important tools for improving patient safety and quality of care, especially by promoting the practice of evidence-based medicine.

Despite this potential for quality improvement, however, few physician practices use EPRs. Nevertheless, interest in EPRs is substantial. A recent survey in 2005 indicated that among 1061 respondents to a random sampling of members of the Medical Group Management Association in US, one in five said they were using an EPR and 40 percent of those without one told they plan to acquire the technology within the next two year. Clearly, the EPR is of growing importance for many physician practices.

In face of adoption barriers, there has been much research outlining the healthcare system’s move towards EPRs for example (Ross and Lin, 2003; Tachinardi et al. 2001; Van’t Riet et al., 2001). However, most of these studies are US based. This study attempts to investigate the facilitating and inhibiting factors that affect health care practitioners to adopt EPRs in Hong Kong; to understand the health care practitioners on their attitude and knowledge towards EPRs in health care practice and to explore the existing utilization and future intention on EPRs in private sector of the health care industry.

2 THEORETICAL FRAMEWORK

To predict acceptance of technology, a number of intention-based theories have evolved, i.e. the theory of planned behavior (Ajzen, 1991), theory of reasoned action (Fishbein and Ajzen, 1975) and the technology acceptance model (Davis, 1989; Davis et al. 1989). According to these theories, user beliefs and attitudes about IT influence adoption and usage behaviours. With few exceptions, however, most studies using these theories have ignored the temporal dimension and the antecedent variables...
that may affect beliefs and attitudes at different stages of the adoption process. Those that have, stress its importance and the need for further study. For example, Venkatesh and Davis (2000) report that the same variables had different effects at different stages of the adoption process, and Fichman and Kemerer (1999) emphasize the need to capture the time of deployment instead of, or in addition to, time of acquisition as the bases for diffusion modeling, driven the observed pattern of cumulative adoptions varies depending on which event in the assimilation process (i.e. acquisition or deployment) is treated as the adoption event. Further, Agarwal and Prasad (1997) support this view that intention-based models may not explain user adoption behavior at the different stages of the adoption process.

Based on this evidence, the current study considers the Rogers’ (1995) stage-based diffusion of innovation model to be the most appropriate to guide its investigation of the formation and change over time of user attitudes and subsequent acquisition and deployment decisions.

Everett Rogers defines diffusion as “the process by which an innovation is communicated through certain channels over time among the members of a social system” (Rogers, 1983, p.5) where innovation has been described as an idea, material, or artifact perceived to be new by the relevant unit of adoption (Zaltman, Duncan, and Holbek, 1973). There are two types of communication channels have been influential in diffusing technology – mass media channels and interpersonal channels. Mass media are radio, television, newspapers, and so on, which enable a source of one or a few individuals to reach an audience of many. And interpersonal channels are face to face, telephone, and personal networks. In his review of innovation diffusion, Rogers (1995) reported mass media channels were most influential in introducing potential adopters to an innovation, whereas interpersonal channels were more influential in subsequent stages.

Innovation diffusion research postulates that many different outcomes are of interest in technology adoption, including the initial adoption, the subsequent routinization and infusion of the innovation. This view is consistent with the stage model as proposed and empirically validated by Cooper and Zmud (1990). These stages of implementation (as shown below) are not necessarily sequential, and should be considered activities that may occur in parallel (Cooper & Zmud 1990):

- **Initiation** – Analyzing organizational needs and potential IT solutions
- **Adoption** – Negotiating to get organizational backing for IT implementation
- **Adaptation** – Developing, installing and maintaining the IT application, revising/developing organizational procedures, training of end-users
- **Acceptance** – Inducing the organizational members to use the technology
- **Routinization** – Encouraging the use of the IT application as a normal activity
- **Infusion** – Effective use of the technology results in the intended benefits (increased organizational effectiveness) of the IT being obtained.

Initiation, adoption and adaptation require both managerial and end-user input and buy-in, and the remaining three stages require necessary dialogues between organizational members for progression through each stage to occur. Additionally, it is widely recognized that successful implementation depends upon gaining organizational members’, targeted as end-users of the innovation, appropriate and committed use of an innovation (Leonard-Barton and Deschamps 1988; Klein and Sorra 1996). It is through the development of a critical mass of individual routinization and infusion that eventual organizational infusion of an innovation is achieved (Tornatzky and Fleischer 1990; Klein and Sorra 1996), and organizational benefits might then be obtained.

Based on the situation in Hong Kong’s clinical practices, most private clinics are either solo practices or partnerships of a few medical doctors that are small in size. Thus the respective process on EPR initiation, adaptation and acceptance are rather straightforward. In this regard, we would like to focus our study on the other three different stages: adoption, routinization and infusion on EPRs in supporting medical practice.

The measure on adoption is based on whether the organization has implemented any EPR. Routinization is measured by the usage of the EPR according to the daily tasks of a clinic. Infusion is measured by the extent of the EPR being integrated with other internal systems within the clinic or external systems outside the clinic. Moreover, the antecedents on these three essential stages will be identified.
3 DATA COLLECTION

A trivial approach to study different stages of a clinic would be longitudinal tracing, but this takes years on the data collection. Nevertheless, we assume different clinics have different extents on the diffusion stages. Some earlier adopters would be more focused on infusion at current moment while some late adopters are still struggling with the routinization of the EPR practice. Thus a cross sectional approach would be adequate for understanding on the diffusion situation of EPRs in Hong Kong. Moreover, we suppose most clinics are not purely on a single stage of the EPR diffusion, they would be 70% adopting on the EPR, with 50% routinizing the EPR practice and 5% infusing the EPR with their daily tasks. In this regard, cross-sectional approach would make more sense to have the overall picture.

Data for this study will be firstly gathered in a focus group interview from which the possible antecedents and measurements on adoption, routinization, and infusion will be determined. A structured questionnaire will be constructed based on the literature with amendments from the focus group to fit the Hong Kong medical practice. The questionnaire will be pilot tested, revised if necessary, and then sent extensively by mail to private sector doctors. From the survey, the respondent (the principle doctor who hosts the clinic) will be asked about the current usage on EPR which consists of the three dimensions on the extent of adoption, routinization, and infusion. The respondent will then evaluate the importance of some pre-defined antecedents on the three respective dimensions of usage. Some open-ended questions will be supplemented any other antecedents not to be included.

The Hong Kong Medical Association (HKMA) has available to the public on the Internet a directory of its members by clinical specialty and geographic location. This list includes essentially all registered private practice doctors in Hong Kong, estimated at about 4,000 in total.

As the data collection instrument will be a mailed questionnaire, we will randomly sample 2,000 doctors listed in the HKMA directory. For the replied respondent, we will mail a $50 Park’N shop couple (for the first 400 respondents) and a summary of our finding as a reward.

4 METHODOLOGY

We shall obtain the means, standard deviations, and bivariate correlations for all data used to analyze predictions of all variables. We intend to perform a factor analysis on the reasons for adopting new technologies as well as those for not adopting these technologies. A multiple ordinary least squares (OLS) regression analysis will be the primary statistical technique to be employed in our study. We shall control for complementarities in the variables and also check whether control variables will have any significant influence on the data.

5 CONCLUSIONS

With our results on the survey, we will understand the facilitating and inhibiting factors as well as the existing practice of EPRs in Hong Kong private clinic. This study will establish a theoretical examination on the diffusion model and inject the managerial insight on how to utilize the EPRs to a greater extent.

ACKNOWLEDGEMENTS

This research was supported in part by The Hong Kong Polytechnic University under grant number A-PA7X.

REFERENCES

Agarwal, R., Prasad, J., 1997. The role of innovation characteristics and perceived voluntariness in the acceptance of information technologies, Decision Sciences, 28, 3, 557 – 582.


MANAGEMENT INFORMATION SYSTEMS IN ROMANIAN UNIVERSITIES

Ana-Ramona Bologa, Razvan Bologa
Computer Science Department, Academy of Economic Studies, Piata Romana 6, Bucharest, Romania
ramona.bologa@ie.ase.ro, razvanbologa@ase.ro

Gheorghe Sabau, Mihaela Muntean
Computer Science Department, Academy of Economic Studies, Piata Romana 6, Bucharest, Romania
sabau@ase.ro, munteanm@inforec.ase.ro

Keywords: The paper Romanian universities, management information systems, ERP systems, integrated information solutions.

Abstract: This paper shortly presents the situation of the Romanian universities regarding information systems implementation and deployment. The information presented is the result of a study regarding the current state of the Romanian universities in the process of data and information system integration, performed at the end of 2007 in 35 accredited universities. This study was used as a base for identifying and analyzing the main factors of influence for developing an integrated university environment and for identifying concrete action directions for accomplishing that integration.

1 INTRODUCTION

The implementation of an information system dedicated to the university management is nowadays a fundamental option for the greatest majority of the universities that understood the new trends at international level. A higher-education information system can be used as a vehicle for professionalizing and transforming the traditional universities and for developing an integrated and standardized Romanian higher education environment, in the European context.

The present paper will present the results of a first phase of a national research project, “Integrated Information Solutions for Competitive Management in Romanian Universities”. The objective of this first phase was to perform a study of Romanian universities in order to obtain a complete view over the present situation of existing management information systems, their integration level and the problems faced in using those systems.

The following phases will analyse the solutions that foreign universities apply, and finally make a pilot implementation in the Academy of Economic Studies of Bucharest.

2 PROBLEM FORMULATION AND GENERAL CONCEPTS

Identifying the most important specific aspects related to the implementation of university governance systems in Romanian, and generally speaking Eastern European, universities is the focus of our current research. Local universities have different process from their Western counterparts and that is why implementations of solutions developed in Western countries had limited success.

The incompatibility with international solutions has led to a situation where local universities are using a high number of various small and poor information systems developed in-house or by small local companies. Almost every faculty or department has its own software applications, developed in-house, applications that use various operation systems, tools, databases and protocols. Those applications are managed by that specific faculty/department and there is no integrated view over the activities developed in the university.

Our team has achieved a study about the current state of the Romanian universities in the process of data and information system integration.

This study is going to be used as a base for identifying and analyzing the main factors of
influence for developing an integrated university environment and for establishing concrete action directions for accomplishing that integration.

We used as a starting point the list of all Romanian universities, offered by the Ministry of Education (Romanian universities), which consisted of 56 accredited state universities, 32 accredited private universities and 24 temporary accredited private universities.

3 APPLICATION INTEGRATION LEVELS

Through our research we will use the conventional three-tier approach. When trying to connect to a certain application, its architecture must be considered. Most applications have a three layer structured architecture (Microsoft - Integration Patterns): Presentation level – is the level that displays the information for the final user and allows him to input data; Business logic level – contains the business functions that action on business data; Data level – accomplishes the persistent data storing in data stores. This level is also called resource level.

Similarly, there are three connection ways between applications and integration level (Microsoft - Integration Patterns): Presentation level integration – the integration level can extract information from the user presentation level using a technique called “screen scraping”; Function level integration – the interaction between the integration level and the business logic level is accomplished by application or service interfaces; Data level integration – the integration level can move data to and from data level.

4 SOFTWARE SOLUTIONS FOR HIGHER EDUCATION MANAGEMENT ON ROMANIAN MARKET

The first phase of our research was to gather data about the current commercial systems used by universities. As expected, these were marginal applications developed by local companies with a limited presence. Our field studies revealed that the number of players on the market was reduced, the following being the most representative:

SICOB (Information System for Public Organization Management) should be a result of a project financed in 2001 by the Ministry of Education in a first attempt of covering the gaps the Romanian universities presented in the field of applications for institution management. It was developed by a Romanian software company and proposed three modules: Financial, Research and Payments. This solution was implemented in about 40 public universities and some of them are still using it.

University Management System (UMS) should be an integrated information system dedicated to higher education institutions (public or private). Developed and provided by a Romanian company, Redpoint SA since 2004, UMS ensures the management of academic processes, students, academic staff, tuition fees, admission process, graduation process, scholarships, lodging and accommodation, diplomas, etc.

Naum Consult System was developed by Naum Consult SRL and has been provided since 2003. The system has modules for: accounting, provision management, assets management, student management, HR, financial.

GESCO 2001 should be an information system for the management of higher education institutions schooling process developed by a Romanian company, Genisoft Group SA. The system offers many facilities such as: student performances management, budget assessment at department, course and allows a correlation with data from SICOB accounting module.

5 STUDY RESULTS ANALYSIS

5.1 Implemented Solutions Analysis

The study was conducted on a number of 35 of the Romanian universities, including both public and private institutions and the full range of Carnegie segments. This can be considered a representative sample of Romanian universities, encountering a percent of 31.25% of total accredited universities.

This study addressed a number of questions such as:
- What kinds of information systems were developed within Romanian universities and for what functional areas?
- Do those information systems succeed in fulfilling the information requests of Romanian universities?
- Which were the major suppliers of software solutions for Higher Education Management on Romanian market?

Almost all universities have a financial solution, and SICOB was the major vendor for the financial
module, encountering 36% of total number of implementations, while Naum Consult follows with a percent of 12%. Anyway, most universities use self-developed solutions for financial and accounting management activities.

Most universities felt the need to implement a Students management module, so that 77% of universities currently use such a module. 41% of those preferred the UMS solution, 11% chose Gesco, 11% chose Naum Consult and there is an important percent of 37% that still use self developed solutions. Although the Student module it wasn’t very popular two or three years ago, the universities have caught up and now most of them use a Student solution, even if not a professional one.

Self-developed solutions are very popular for Financial and HR/Payrolls modules, as they were preferred in order to make cost savings but also to cover the incomplete functionality of the existing software. Another important remark is that big universities use all the HE ERP modules, commercial or self developed. Small universities, on the other side don’t use commercial solutions, because of their cost. Public universities always use Financial and HR/Payrolls modules, regardless of developer, as they have to make mandatory specific reports. Our study indicates that the number of players on the HE ERPs market is small. Most universities use in-house or low-end solutions.

### 5.2 Integration Level

A real process level integration is presented only by University Management System solution. As its developer, the company Redpoint, initiated collaboration with SAP corporation in order to develop the following modules: financial accounting, management accounting, payroll, virtual library, this solution has a big chance to become a complete and competitive software package for university management. In Naum Consult we have a partly data level integration by using a common database in several different modules. SICOB, on the other side, is a simple accounting solution, so the integration can be made only external, by another application or support platform. As for the university in-house solutions, there the integration is almost inexistent. The modules doesn’t communicate between them, at most they are exported from one module, processes and transformed and then imported into other modules.

### 5.3 The Example of the Academy of Economic Studies, Bucharest

A preliminary analysis of the information system in our university, the Academy of Economic Studies, revealed the existence of 7 major independent information systems, which are independent and communicate with each other only by export files or specially developed transfer programs. Here is a short description of those systems and their relationships with each other.

1. **Student Management Application** is a client-server application that uses a Microsoft SQL Server 2005 database. Its interface was developed in Visual Fox Pro and Visual Basic. It includes three modules:
   a. **Student Master Data**: includes student information, faculties, degrees and programs information, their curricula, courses etc. It uses a specially developed transfer program in order to extract information from Admission Management Application.
   b. **Student Academic History**: includes a list of all grades up to and including current term, earned credit points, reprogrammed exams. The secretaries are responsible for updates.
   c. **Student Bills**: includes all kind of student fees, charges, payments. Data are updated by the pay offices (approximately 10 workstations).

2. **Student Admission Management Application** uses as inputs data regarding student personal data, their admission options, faculties and specializations, student photos etc. It provides basic

---

Table 5.1: University application analysis by architecture and implementation technologies.

<table>
<thead>
<tr>
<th>Operating system</th>
<th>Guesco</th>
<th>UMS</th>
<th>Sicob</th>
</tr>
</thead>
<tbody>
<tr>
<td>Operating system</td>
<td>Windows 2000 Server</td>
<td>Windows 2003, Unix, Linux</td>
<td>Sun Server</td>
</tr>
<tr>
<td>Database</td>
<td>SQL Server</td>
<td>Oracle SQL Server</td>
<td>Progress Workgroup DB</td>
</tr>
<tr>
<td>Development tools</td>
<td>Microsoft.NET (ASP, Web, XML, C)</td>
<td>Java J2EE, SWING, JSP, Struts, EJB</td>
<td>Progress Actuate Developer</td>
</tr>
<tr>
<td>Reporting and analysis</td>
<td>Embedded, based on XML, self transfer in .pdf or .xls</td>
<td>Embedded, integrated with MS Office</td>
<td></td>
</tr>
</tbody>
</table>
data for the Student Management Application and it uses outputs from Student Bills module, in order to check out tuition fees payment. This application was developed entirely in Visual Fox Pro.

3. The Academy Web Site: centralizes all the relevant information for various visitors category, students, inside and outside the university. Data are not automatically displayed by interrogating a database, but manually filled in by secretaries and then converted to HTML format. As a result, there is a high incidence of human errors in data and many out of date or incomplete on-line information.

4. Social Information System was developed in Visual Fox Pro and includes four modules: Lodging and accommodations-student lodging on university campus; Lodging Fees; Scholarships and financial aids; Transportation reimbursement-public transportation expenses reimbursement. It imports data from the Student Management Application.

5. SICOB is used for financial and accountability specific operation.

6. Human Resources and Payrolls Application uses Microsoft SQL Server database and includes 2 specific modules: Payrolls and Human resources.

7. Library for Universal (L4U): a classic system for library management that uses data exported from Student Management Application and personnel database, after a previous filtering and transformation by a specially designed program.

There is also Web-based Virtual Campus management application for distance learning programs.

6 CONCLUSIONS

The Romanian learning system was recently integrated, at least at a formal level, in the European learning system, automatically bringing the need of adapting to European practices (e.g. ECTS).

The capacity of Romanian universities information systems to respond to the demands and the challenges of European education system was very modest until now. Universities should realize that they are not so drastically different and given that, they should collaborate in order to influence vendors to become more sensitive to higher education needs. This way, ERP vendors should provide some best practice models to reduce the costs of ERP implementation.

European integration of Romanian higher education system might benefit from the support of European funds. But, in order to access the European funds, the universities should manifest interest, make congruent efforts and put some pressure on the decision factors, on one side, and, on the other side, the political support is also very important. For the moment none of them seems to be strong enough.

REFERENCES

Marti Harris, Michael Zastrocky, Jan-Martin Lowendahl , Magic Quadrant for Higher Education Administrative Suites 2006, Gartner Industry Research 16 September 2006;
Ion Lungu, Ana-Ramona Bologa, Adela Bâra, Vlad Diaconita – Integrarea sistemelor informatice, Editura ASE, Bucharest, 2008;
The list of accredited Romanian universities, http://www.edu.ro;
A FEATURE EXTRACTING METHOD FOR TAMPER DETECTION IN PRINTED DOCUMENTS

Yoshiyasu Takahashi, Takaaki Yamada and Seiichi Susaki
Systems Development Laboratory, Hitachi Ltd., Totsuka-ku, Yokohama, Kanagawa, Japan
{yoshiyasu.takahashi.gq, takaaki.yamada.tr, seiichi.susaki.gw}@hitachi.com

Keywords: Feature Extracting Method, Dot-Pattern Code, Digital Watermarking, Electronic Delivery, e-Government.

Abstract: In this paper, we propose our feature extracting method for tamper detection in printed documents. To detect the tamper of the printed document, a feature extracting method is needed. In this paper, we describe our feature extracting method. Our feature extracting method is based on the location of the mean point of each dot. We have estimated our feature extracting method's probability of the collision, its uniformity of the distribution, its invariability during D/A and A/D transform and its invariability during the ordinal change of paper. We have found that our feature extracting method can extract desirable feature value.

1 INTRODUCTION

Since 2000, the Japanese Government has promoted the IT strategy along with the implementation of various structural reforms. (IT Strategic Headquarters, 2007).

However, several issues still remain to solve. One of the issues is the lower utilization rate of e-government.

We consider the reasons of the lower utilization rate are three; complicated preparation for online application, security worry and the lack of the means of the online delivery.

To resolve the third reason, we can envisage a system that the government delivers certificates online and the user also receives them online, and the user prints them with their home printers and use or submit them to the destination in paper format.

To bring the scenario to fruition, the two means are needed; a means to prove that the paper is original print and not copy, and a means to prove that the content of the paper is correct and not tampered.

To detect the copy, the “copy indicator” technique is widely known.

To detect the tamper of the paper, the extracting method of the paper’s feature is needed. With that method, we can extract the feature of the printing document and print it within the paper. After the paper is distributed, the verifier can extract both the feature of the paper and the feature embedded, and compare them. However, the extracting method of the paper’s feature is not well known.

In this paper, we describe the requirements for the feature extracting method and introduce our innovative method to extract the paper’s feature. We also make estimation of our method and compare with another extracting method.

2 REQUIREMENTS FOR THE FEATURE EXTRACTING METHOD

For the feature extracting, the requirements are following five.

1) The feature values calculated from two different documents should not be equal
2) The distribution of the feature values should be uniform.
3) The document should not be calculated from its feature value.
4) The feature values should not vary during the D/A and A/D transform.
5) The feature values should not vary if the paper document is folded or got wrinkles or tainted.

For the feature extracting, the problem is more difficult than the hash values because of the requirement 4 and 5. Because the D/A and A/D transform is inevitable, the feature value should not vary during them. Moreover, because paper is used, folding, wrinkles or taint easily occur, the feature value should keep the same value during the ordinal alteration of the document. At the same time, if the significant change occurs in the document, even if
the change is minute, the feature value should change the other values to represent the change of the document.

3 RELATED WORKS

The ideal feature value has not been found yet. However, several techniques that meet a part of the requirements are studied to extract the paper’s feature to detect the tampering.

There are three types of the paper’s feature below.

1) using the document itself
2) using the semi-fragile watermarking
3) using the feature of the paper

First, there are methods to use (a part of) the original document, such as the text data or the image data. To embed them, 2D-code is usually used. To check the document’s tamper, usually a man power is needed; a verifier scans the code and extract the document’s text data or image data, and compare them, watching side-by-side. Its cons are its easiness to implement and its robustness for the paper’s conditions, and its pros are that it needs a verifier’s help and time to verify.

Second, the use of semi-fragile watermark method has been proposed. The semi-fragile watermark is a watermark which is robust to some degradation such as compression, while it is at the same time destroyed if the embedded image is tampered with. However, the semi-fragile watermark which is robust to the paper’s folding or wrinkles or tainting is not well-known. Therefore, workable proposal of this type of method has not been yet.

Third, using the feature of the paper has been proposed. Suzuki and et al. has proposed the watermarking technique for printed documents by superposing dot pattern blocks on backgrounds of the document image (Suzuki, 2003).

4 PROPOSING THE FEATURE EXTRACTING METHOD

4.1 Brief Overview

Figure 1 shows the overview of our document verification system.

To print the certificate, the following procedures are performed.

1) The issuer makes the certificate.
2) The system calculates the feature value from the certificate, and embeds it into the certificate itself. The concrete means of extracting and embedding of the feature value are described later.
3) The system prints the feature value embedded certificate with printers. In the electronic delivery, this process can be performed at the user’s home.

After the issue, because the certificate circulates among many people, there is a risk of tampering the certificate. Therefore, the submitted certificate needs to be verified. With the following procedures, the verifier can verify the certificate.

1) The verifier scans the certificate and detect the embedded feature value.
2) The verifier also calculates the feature value from the submitted certificate.
3) The verifier compares the two feature values. When the significant difference is detected, the verifier judges the certificate as tampered. The concrete means of extracting and embedding the feature value are described below.
4.2 Extracting the Feature Value

In our method, we use the median point as the feature value of the document, as shown in Figure 2. In the Figure 2, the plus sign illustrates the median point of the character “4”.

![Figure 2: Sample of the feature value.](image)

The benefits of the median point as the feature value are the robustness against the blur because of the scattering of the tonner or ink of printers and of the focus error of scanners. Figure 2 also illustrates this merit. In Figure 2 left, the character “4” is not blurred. This illustrates the character “4” to calculate the feature value in issuer’s computer. After the printing, circulating, and scanning, the character may blur. The Figure 2 right illustrates this situation.

Although the blur, the median point of the character “4” of the Figure 2 right does not vary very much. That is because the width of the line of the character is almost the same.

To embed the feature value into the Document, we use the dot-pattern code (Takahashi 2007, 2008), which the authors have developed before. The dot-pattern code is the information embedding method onto the paper with minute dots.

5 ESTIMATION

5.1 Assessment Viewpoints of the Feature Value

The viewpoints of the assessment are following five.
(A) probability of the collision
(B) uniformity of distribution
(C) probability of inverse calculation
(D) invariability during D/A and A/D transform
(E) invariability during the ordinal change of paper

5.2 Probability of the Collision and Uniformity of Distribution

To estimate the probability of the collision and the uniformity of distribution, we calculated the median points of characters. The characters are alphabet and number of Times New Roman font. We drew a character one by one on the white image and calculated its median point.

The result is shown in Figure 3. In Figure 3, the median points of each character are plotted with dots. From the Figure 3, though there is some bias, we can find that the median points are distributed almost uniformly. Moreover, there are no dots which collide against another dot.

To compare, we also calculated the area of each alphabet and number. In Figure 4, the histogram of the result is shown. The x-axis is the area and the y-axis is occurrence. From Figure 4, we can find the area is distributed not uniformly. Indeed, nine characters collided; i.e. they have the same area. Therefore we can conclude that our feature value is better than the area.

![Figure 3: Median points of a character.](image)

![Figure 4: Area of a character.](image)

5.3 Invariability during D/A and A/D Transform

To estimate the invariability during D/A and A/D transform, we viewed the difference of the median point’s change between before and after printing the document.
We calculate the median point and area of a character “a” in digital format. Then, we print it and scanned and calculated the median point and area again from the scanned image. Before calculating the median point and area, we converted the scanned image to binary image with various thresholds.

Table 1: Invariability during D/A and A/D transform.

<table>
<thead>
<tr>
<th>thrs</th>
<th>median point</th>
<th>area diff</th>
</tr>
</thead>
<tbody>
<tr>
<td>192 (149.12, 146.86)</td>
<td>2.33</td>
<td>260 128</td>
</tr>
<tr>
<td>128 (150.28, 148.44)</td>
<td>0.58</td>
<td>186 54</td>
</tr>
<tr>
<td>64 (150.02, 148.22)</td>
<td>0.76</td>
<td>152 20</td>
</tr>
<tr>
<td>32 (150.49, 148.31)</td>
<td>0.78</td>
<td>116 16</td>
</tr>
<tr>
<td>16 (150.41, 148.89)</td>
<td>1.34</td>
<td>108 24</td>
</tr>
<tr>
<td>8 (150.46, 148.75)</td>
<td>0.44</td>
<td>104 28</td>
</tr>
<tr>
<td>orig (150.08, 148.98)</td>
<td>---</td>
<td>132 ---</td>
</tr>
</tbody>
</table>

The result is shown in Table 1. From Table 1, we can find that though the threshold varies (thrs), the median point does not vary very much. On the other hand, the area varies very much and the difference of the area of the original image. There are twelve characters whose area’s difference is less than 54. Therefore we cannot distinguish these twelve characters with area.

5.4 Invariability during the Ordinal Change of Paper

To estimate the invariability during the ordinal change of paper, we viewed the difference of the median points’ change before and after folding of the paper. The experiment is done as 5.3.

Table 2: Invariability during the ordinal change of paper.

<table>
<thead>
<tr>
<th>thrs</th>
<th>median point</th>
<th>area diff</th>
</tr>
</thead>
<tbody>
<tr>
<td>192 (146.03, 144.40)</td>
<td>6.11</td>
<td>296 164</td>
</tr>
<tr>
<td>128 (150.59, 148.79)</td>
<td>0.53</td>
<td>164 32</td>
</tr>
<tr>
<td>64 (150.51, 148.59)</td>
<td>0.57</td>
<td>59 73</td>
</tr>
<tr>
<td>orig (150.08, 148.98)</td>
<td>---</td>
<td>132 ---</td>
</tr>
</tbody>
</table>

The result is shown in Table 2. In this experiment, our median point shows near to original median point. Therefore we can conclude that our proposed method has good invariability during the folding of the paper.

6 CONCLUSIONS

In this paper, we have introduced our feature extracting method of paper document. Our feature extracting method is based on the location of the mean point of each dot, and is expected to be applicable to home printers such as inkjets.

We have estimated the probability of the collision and the uniformity of the distribution of our feature extracting method. We have found that the feature value extracted from one character is distributed uniformly and do not collide each other.

We have also checked invariability during D/A and A/D transform and we have found that almost every character can be distinguished with our feature extracting method even after D/A and A/D transform. We have also found that our feature extracting method is better than the area.

We have also checked the invariability during the ordinal change of paper and found that our method has enough and better invariability than the area. Therefore, we can conclude our proposed method can extract desirable feature value.

REFERENCES


THE AFFECTION OF DEMOGRAPHIC CHARACTERISTICS ON MEDIA CHOICE

A Case about Chinese Miniature Automobile Consumers

Dao-ping Chen
School of Economy and Business Administration, Chongqing University, Chongqing, P.R. China
School of Economics and Management, Chongqing Normal University, Chongqing, P.R. China
cdp8cdp@163.com

Wei Liu
School of Economy and Business Administration, Chongqing University, Chongqing, P.R. China
liuwei@cqu.edu.cn

Keywords: Miniature automobile consumers, demographic characteristics, media choice, optimal scaling regression.

Abstract: This paper explores the media choice of Chinese miniature automobile consumers in obtaining purchasing information about miniature automobile. A worldwide survey involved a majority of Chinese areas was conducted. The survey focused on consumers' demographic characteristics and media choice. The result of the survey shows that consumers chiefly choose newspaper, TV or automobile exhibitions as the media to obtain purchasing information about miniature automobile. Then, based on optimal scaling regression analysis of statistics, the influences of demographic characteristics on media choice were studied. The study finds that consumers' demographic characteristics significantly affect their media choice. In detail, consumers' living city, occupation and education are three most important characteristics that affect the media choice of consumers.

1 INTRODUCTION

Consumers can obtain purchasing information from various media such as newspaper, TV, magazine, or Internet. Many scholars studied the consumers' media choice behaviour (Wei and Pan, 1999, Hung, Gu, and Tse, 2005, Geist, 2004, Ducoffe, 1996, Yoon and Kim, 2001, Milsom, 2003). These studies found that consumers which are from different countries or purchase different products have different choice to media in obtaining purchasing information.

There are many factors influencing consumers' media choice. Demographic characteristics, such as age, education, occupation, and income, are generally regarded as important factors affected the media choice of consumers (Westbrook and Fornell, 1979, Andreasen and Ralchford, 1976, Schaninger and Sciglimpagia, 1981, Newman and Staelin, 1973, Chiteji and Stafford, 1999, Claxton, Fry, and Portis, 1974). This study has the main objective of empirically determining the affection of demographic characteristics on media choice of Chinese miniature automobile consumers.

2 QUESTIONNAIRE AND RESPONSE

Newspaper, magazine, TV, Internet, and automobile exhibition are selected as the resources from which Chinese consumers get purchasing information about miniature automobile. Gender, living city, occupation, education, age, and monthly household income are used to describe demographic characteristics. Gender has two levels (1=male, 2=female), living city four levels (1=big city, 2=city of middling size, 3=county seat, 4=villages and towns), occupation seven levels (1=government servant, 2=employee of national enterprise, 3=employee of private enterprise, 4=employer of individual enterprise or partnership enterprise, 5=farmer, 6=professional (lawyer, accountant, teacher, doctor, athlete, reporter etc.), 7=other),
education six levels (1=junior high school or below, 2=senior high school, 3=technical secondary school, 4=junior college, 5=college or university, 6=graduate student), age seven levels (1=18-21 years, 2=22-25, 3=26-29, 4=30-34, 5=35-39, 6=40-59, 7=60 or above 60 years), and monthly household income four levels (1=less than ¥2 000, 2=¥2 000 to less than ¥5 000, 3=¥5 000 to less than ¥8 000, 4=more than ¥8 000).

The data used in this study are obtained from a large survey sponsored by an automobile group company in China. 2 630 questionnaires are sent to 263 dealers of the automobile group through mails, commencing in December 2003, at three month intervals. A dealer selected is responsible for 10 questionnaires. The objects of the survey are consumers who visit the shop of dealers and have intention to purchase miniature automobile. The survey areas involve 31 provinces, municipalities and autonomous regions of Chinese Mainland. A total of 2 623 usable responses are received, yielding a response rate of 99%. 280 questionnaires in 2 623 are discarded because they are not perfectly filled. The remaining 2 343 questionnaires are used for the final analysis. All data are processed by SPSS 15.0.

3 RESULTS

3.1 The Media Choice of Chinese Consumers

The result of the media choice by Chinese miniature automobile consumers is shown in table 1. The percentage of the media choice breakdown in descending order is as follows: newspaper (42.1%), TV (36.8%), automobile exhibition (23.9%), Internet (20.2%), and magazine (12.0%). The percentage of newspaper is higher than that of TV, which shows that it is more possible for consumers to obtain advertising information from newspaper than from TV although Chinese families almost have TV. Though Internet has had a rapid development in China in recent years, it is early for Internet to be regard as a main medium of consumers for obtaining automotive advertising information. It appears surprising that the percentage of magazine, as an important traditional medium, is only 12.0%, which may be because Chinese miniature automobile consumers are not known as chief target of magazine.

<table>
<thead>
<tr>
<th>Information sources</th>
<th>Frequency</th>
<th>Percent (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Newspaper</td>
<td>986</td>
<td>42.1</td>
</tr>
<tr>
<td>Magazine</td>
<td>281</td>
<td>12.0</td>
</tr>
<tr>
<td>TV</td>
<td>863</td>
<td>36.8</td>
</tr>
<tr>
<td>Internet</td>
<td>474</td>
<td>20.2</td>
</tr>
<tr>
<td>Automobile exhibition</td>
<td>559</td>
<td>23.9</td>
</tr>
</tbody>
</table>

3.2 The Influence of Demographic Characteristics on the Media Choice

We conducted optimal scaling regression analysis. Newspaper, TV, Internet, automobile exhibition, or magazine is dependent variables and gender, living city, occupation, education, age and monthly household income are independent variables. The regression result is shown in table 2.

3.2.1 Newspaper

From table 2, we know that the regression model is statistically significant at the 0.01 level (F=3.210, Sig.<0.01). Of six demographic characteristics, four characteristics, living city (F=29.476, Sig.<0.01), occupation (F=5.929, Sig.<0.01), monthly household income (F=7.693, Sig.<0.006<0.01), and age (F=2.901, Sig.<0.021<0.05) have a significant influence on dependent variable newspaper. Living city (62.2%) is the most important characteristics, followed by monthly household income (20.0%), occupation (8.2%), and age (5.3%). Gender and education have not significant affection on dependent variable newspaper.

3.2.2 Magazine

From table 2, we can know that the regression model is statistically significant at the 0.01 level (F=2.600, Sig.<0.01). From table 2, we can see that occupation (F=11.032, Sig.<0.01), education (F=19.720, Sig.<0.01) and age (F=4.858, Sig.<0.01) produce significant influence on dependent variable magazine. The three demographic characteristics breakdown in descending order of importance are as follows: education (55.6%), occupation (32.7%), age (11.6%).

From table 2, we can know that gender, living city and monthly household income have not significant affection on choice of consumers on magazine.
### Table 2: Results of Regressions of Consumer Characteristics on Media Choice (N=2,343).

<table>
<thead>
<tr>
<th>Media</th>
<th>Characteristics</th>
<th>ANOVA</th>
<th>Beta</th>
<th>df</th>
<th>F</th>
<th>Sig.</th>
<th>Importance</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Newspaper</strong></td>
<td>Gender</td>
<td>F(16, 2,326) = 2.120</td>
<td>-0.022</td>
<td>1</td>
<td>1.063</td>
<td>.030</td>
<td>.015</td>
</tr>
<tr>
<td></td>
<td>Living city</td>
<td></td>
<td>.114</td>
<td>2</td>
<td>29.476</td>
<td>.000</td>
<td>.622</td>
</tr>
<tr>
<td></td>
<td>Occupation</td>
<td></td>
<td>-0.054</td>
<td>6</td>
<td>5.929</td>
<td>.000</td>
<td>.082</td>
</tr>
<tr>
<td></td>
<td>Education</td>
<td></td>
<td>-0.018</td>
<td>2</td>
<td>0.625</td>
<td>.536</td>
<td>.028</td>
</tr>
<tr>
<td></td>
<td>Age</td>
<td></td>
<td>.036</td>
<td>4</td>
<td>2.901</td>
<td>.021</td>
<td>.053</td>
</tr>
<tr>
<td></td>
<td>Income</td>
<td></td>
<td>-0.059</td>
<td>1</td>
<td>7.933</td>
<td>.006</td>
<td>.200</td>
</tr>
<tr>
<td><strong>Magazine</strong></td>
<td>Gender</td>
<td>F(19, 2,323) = 2.600</td>
<td>-0.033</td>
<td>1</td>
<td>2.405</td>
<td>.121</td>
<td>.013</td>
</tr>
<tr>
<td></td>
<td>Living city</td>
<td></td>
<td>-0.020</td>
<td>2</td>
<td>0.825</td>
<td>.438</td>
<td>-0.024</td>
</tr>
<tr>
<td></td>
<td>Occupation</td>
<td></td>
<td>-0.071</td>
<td>6</td>
<td>11.032</td>
<td>.000</td>
<td>.327</td>
</tr>
<tr>
<td></td>
<td>Education</td>
<td></td>
<td>-0.010</td>
<td>5</td>
<td>19.720</td>
<td>.000</td>
<td>.556</td>
</tr>
<tr>
<td></td>
<td>Age</td>
<td></td>
<td>.046</td>
<td>3</td>
<td>4.858</td>
<td>.002</td>
<td>.116</td>
</tr>
<tr>
<td></td>
<td>Income</td>
<td></td>
<td>-0.008</td>
<td>2</td>
<td>0.144</td>
<td>.866</td>
<td>.012</td>
</tr>
<tr>
<td><strong>TV</strong></td>
<td>Gender</td>
<td>F(15, 2,327) = 2.108</td>
<td>-0.028</td>
<td>1</td>
<td>1.799</td>
<td>.180</td>
<td>.078</td>
</tr>
<tr>
<td></td>
<td>Living city</td>
<td></td>
<td>-0.078</td>
<td>6</td>
<td>13.747</td>
<td>.000</td>
<td>.416</td>
</tr>
<tr>
<td></td>
<td>Occupation</td>
<td></td>
<td>-0.050</td>
<td>6</td>
<td>5.416</td>
<td>.000</td>
<td>.153</td>
</tr>
<tr>
<td></td>
<td>Education</td>
<td></td>
<td>-0.066</td>
<td>4</td>
<td>9.063</td>
<td>.000</td>
<td>.200</td>
</tr>
<tr>
<td></td>
<td>Age</td>
<td></td>
<td>-0.007</td>
<td>2</td>
<td>1.721</td>
<td>.179</td>
<td>.071</td>
</tr>
<tr>
<td></td>
<td>Income</td>
<td></td>
<td>.032</td>
<td>1</td>
<td>2.345</td>
<td>.126</td>
<td>.081</td>
</tr>
<tr>
<td><strong>Internet</strong></td>
<td>Gender</td>
<td>F(23, 2,319) = 16.133</td>
<td>.011</td>
<td>1</td>
<td>3.111</td>
<td>.577</td>
<td>-0.004</td>
</tr>
<tr>
<td></td>
<td>Living city</td>
<td></td>
<td>-0.053</td>
<td>2</td>
<td>6.917</td>
<td>.001</td>
<td>.060</td>
</tr>
<tr>
<td></td>
<td>Occupation</td>
<td></td>
<td>.161</td>
<td>6</td>
<td>52.029</td>
<td>.000</td>
<td>.331</td>
</tr>
<tr>
<td></td>
<td>Education</td>
<td></td>
<td>-0.221</td>
<td>5</td>
<td>91.927</td>
<td>.000</td>
<td>.524</td>
</tr>
<tr>
<td></td>
<td>Age</td>
<td></td>
<td>.084</td>
<td>6</td>
<td>18.557</td>
<td>.000</td>
<td>.062</td>
</tr>
<tr>
<td></td>
<td>Income</td>
<td></td>
<td>-0.047</td>
<td>3</td>
<td>5.598</td>
<td>.001</td>
<td>.027</td>
</tr>
<tr>
<td><strong>Automobile exhibition</strong></td>
<td>Gender</td>
<td>F(16, 2,326) = 2.479</td>
<td>-0.049</td>
<td>1</td>
<td>5.413</td>
<td>.020</td>
<td>.187</td>
</tr>
<tr>
<td></td>
<td>Living city</td>
<td></td>
<td>-0.052</td>
<td>2</td>
<td>5.700</td>
<td>.003</td>
<td>.228</td>
</tr>
<tr>
<td></td>
<td>Occupation</td>
<td></td>
<td>-0.060</td>
<td>6</td>
<td>8.227</td>
<td>.000</td>
<td>.264</td>
</tr>
<tr>
<td></td>
<td>Education</td>
<td></td>
<td>-0.054</td>
<td>4</td>
<td>6.099</td>
<td>.000</td>
<td>.228</td>
</tr>
<tr>
<td></td>
<td>Age</td>
<td></td>
<td>.014</td>
<td>1</td>
<td>0.473</td>
<td>.492</td>
<td>.026</td>
</tr>
<tr>
<td></td>
<td>Income</td>
<td></td>
<td>.047</td>
<td>2</td>
<td>4.972</td>
<td>.007</td>
<td>.068</td>
</tr>
</tbody>
</table>

#### 3.2.3 TV

The result of ANOVA of regression model on TV shows that the model is statistically significant at the 0.01 level ($F=2.108, \text{Sig.}=0.008$). Living city ($F=13.747, \text{Sig.}<0.001$), occupation ($F=5.416, \text{Sig.}<0.001$) and education ($F=9.063, \text{Sig.}<0.001$) produce significant influence on dependent variable TV. The three characteristics are put in descending order of importance as follows: living city (41.6%), education (20.0%), and occupation (15.3%). Moreover, from table 2 we can know that gender, age and monthly household income have not significant affection on dependent variable TV.

#### 3.2.4 Internet

The result of ANOVA of regression model on Internet shows that the model is statistically significant at the 0.01 level ($F=16.133, \text{Sig.}<0.001$). From table 2, we can see that living city ($F=6.917, \text{Sig.}<0.001$), occupation ($F=5.202, \text{Sig.}<0.001$), education ($F=91.927, \text{Sig.}<0.001$), age ($F=18.557, \text{Sig.}<0.001$) and monthly household income ($F=5.598, \text{Sig.}<0.001$) produce significant influence on dependent variable Internet. Owing to table 3, we can know that education (52.4%) is the most important of the five characteristics, followed by occupation (33.1%), age (6.2%), living city (6.0%) and monthly household income (2.7%). Gender is only characteristic which is not statistically significant in this case of Internet.

#### 3.2.5 Automobile Exhibition

The result of ANOVA of regression model shows that the model on automobile exhibition is statistically significant at the 0.01 level ($F=2.479, \text{Sig.}<0.01$). From table 2, we can see that consumers' gender ($F=5.413, \text{Sig.}=0.020<0.05$), living city ($F=5.700, \text{Sig.}=0.003<0.01$), occupation ($F=8.227,$
Sig.<0.001), education (F=6.009, Sig.<0.001) and
monthly household income (F=4.972, Sig.=0.007 <0.01) produce significant influence on their choice
on automobile exhibition. The five characteristics
are ranked in descending order of importance as
follows: occupation (26.4%), education (22.8%),
living city (22.8%), gender (18.7%), and monthly
household income (6.8%). Age has not significant
impact on dependent variable automobile exhibition.

4 SUMMARY AND
CONCLUSIONS

The study focuses on the media choice of Chinese
consumers in obtaining purchasing information
about miniature automobile and on the influence of
their demographic characteristics on the media
choice. The results show that Chinese consumers
mainly choose newspaper, TV, and automobile
exhibitions to obtain purchasing information about
miniature automobile. Chinese consumers appear not
to prefer magazine to gain advertising information
about miniature automobile.

The study finds that the demographic characteristics of Chinese consumers produce a significant impact at the 0.01 level on whether consumers select newspaper, magazine, TV, Internet, or automobile exhibition as a medium of obtaining purchasing information. Of all six characteristics variable, living city, occupation and education are the three most important characteristics which affect the media choice. Gender has not significant influence on almost all media choice of consumers.

The results provide substantial implications for
miniature automobile manufacturers which face advertisement decision. If an automaker intends to select a medium from various media for advertising, newspaper or TV should first be considered. Another implication derivable from this study is that manufacturers should conduct a classification on consumers according to their living city, occupation or education characteristics. Manufacturers should integrate several media to publish advertising information suitable for the consumer categories. For example, if they consider to use newspaper, TV or Internet for advertising, automakers should chiefly focus on the appeals of consumers living in the city of middling size, or individual-enterprise or partnership-enterprise employer doing pioneering work or consumers with a low education level in advertising design of Chinese miniature automobile.

This study is somewhat limited in the sense that
the sample is confined to consumers entering the
shop of dealer and having intention to purchase
miniature automobile, therefore, our results are only
a first step toward understanding Chinese miniature
automobile consumers. A more comprehensive
survey that focuses on all consumers of automobile
in China should be conducted in the future. If so, a
more representative sample could be obtained. By
the new sample, a contrast could be done on the
media choice between consumers of miniature
automobile and consumers of other automobile
categories.

REFERENCES

Andreasen, AR., Ralchford, BT., 1976. Factors affecting
consumers’ use of information sources. Journal of
Chiteji, NS., Stafford, FP., 1999. Portfolio choices of
parents and their children as young adults: asset
accumulation by African-American families. American
Claxton, JD., Fry, JN., Portis, B., 1974. A taxonomy of
prepurchase information gathering patterns. Journal of
Consumer Research, 1(3): 35-42.
Ducoffe, RH., 1996. Advertising Value and Advertising on
the Web. Journal of Advertising Research, 36
(September/ October), 21-35.
Issue 6116, p26B-26E, 3p.
Hung, K., Gu, FF., and Tse, DK., 2005. Improving Media
Decisions in China. Journal of Advertising 34
(Spring), p49, 15p.
Research. New Media Age, 8/21/2003, p13, 1/2p, 1c.
Newman, JW., Staelin, R., 1973. Information sources of
durable goods. Journal of Advertising Research, 13(2):
19-29.
Schaninger, CM., Sciglimpagia, D., 1981. The influence
of cognitive personality traits and demographics on
consumer information acquisition. Journal of
Consumer Research, 8(2): 208-216.
Westbrook, RA., Fornell, C., 1979. Patterns of information
source usage among durable goods buyers. Journal of
Marketing Research, 16 (3): 303-312.
Wei, R., Pan, ZD., 1999. Mass Media and Consumerist
Values in the People's Republic of China. International
Journal of Public Opinion Research, 11
(1), 75—97.
Yoon, SJ., Kim, JH., 2001. Is the Internet More Effective
Than Traditional Media? Factors Affecting the Choice
of Media. Journal of Advertising Research, 41
(November/December), 53-60.
WEB BASED COLLABORATIVE DOCUMENT CREATION AND REVIEW SYSTEM

Marius Ioan Podean, Raluca Arba
Faculty of Economics and Business Administration
Babes Bolyai University of Cluj Napoca, Teodor Mihali 58-60, Cluj Napoca, Romania
marius.podean@econ.ubbcluj.ro, raluca.arba@econ.ubbcluj.ro

Loredana Muresan
Faculty of Economics and Business Administration
Babes Bolyai University of Cluj Napoca, Teodor Mihali 58-60, Cluj Napoca, Romania
loredana.muresan@econ.ubbcluj.ro

Keywords: Collaborative document editing, document management, workflow, virtual teams, XML.

Abstract: An important aspect in distributed teams and organization is their ability to manage documents. A collaborative document editing system that integrates functionalities from document management systems, workflow, collaborative editing with support for virtual teams can increase team efficiency and allow users to concentrate their efforts on content development. This paper reports a case study implementing this approach in collaboratively creating scientific papers. The use of XML when treating documents proves to be the appropriate solution to develop user and document centered systems.

1 INTRODUCTION

Creating scientific papers is most of the time a very complex and elaborate task. This type of documents usually is written by more then just one person, and often the authors are located in different parts of the world. This collaborative process is sometimes slowed down by the fact that dealing with the technology used to support it is very much time consuming. Taking in consideration the authors need for location and time independence, the usage of different operating systems and applications, at this moment technology offers solutions that satisfy just partially the aforementioned requirements.

Document management systems (DMS) focus on tracking and storing documents created and exchanged by there authors (Aversano et al. 2001). They provide components for defining metadata for the documents (i.e. date of creation, authors, version etc.), indexing (usually based on metadata), storage and retrieval (based on the unique document identifier).

Workflow management systems (WMS) allow the automation of processes within an organization, enabling greater coordination and control among geographically distributed teams (Nallaparaju et al. 2005). Using WMS, the organization can integrate different software technologies, leading to the improvement of the collaborative activities (Aversano et al. 2001). This technology leaves authors with the consumption of a great deal of time.

Collaborative editing allows multiple persons to edit simultaneously the same document, see who is working on the document and watch in real time the changes that they have made (Raikundalia and Zhang 2004). In order to have sufficient knowledge about the changes that others perform upon the document, group awareness mechanisms have been created, such as: telepointers (multiple cursors of users appear within the document), radar views, multi-user scrollbars and, as shown by Raikundalia and Zhang (2004), structure-based multi-page view, point jumping mechanism and user info list. The main downside of this solution is that the implementations available are platform-specific and can generate conflicts between members when someone changes often content created by others. The collaboratively edited file is stored on the document owner’s computer, leading towards versioning problems when participating members make local copies of the document.
Wikis represent “a piece of software that allows users to add, modify, and/or delete information from a knowledge base via the web” according to Spek (2008). As the author of this definition emphasis as a main characteristic, wikis are anarchistic systems, many implementations allowing anonymous users to modify the content. On wiki systems “conflicts can quickly result in ‘edit wars’ when multiple users keep on reverting each other’s changes because they don’t agree” as shown by Spek (2008).

Creating research papers requires a system that is simple to use and allows users to focus their efforts on the content rather than on the technology used to create it. The system has to support the space independence of the authors and integrate their collaborative effort in a common workplace in order to obtain greater team efficiency (Guerrero et al. 2004). Collaborative work involves information exchange in order to support negotiation and communication between group members and different mechanisms through which the team can regulate and manage itself in order to be goal directed (Millward and Kyriakidou 2004). A more supportive system would have document management facilities and support for task automation. To achieve greater efficiency the system will have to be user centered and non-restrictive regarding de operating system.

In this paper, we start by presenting current technologies used for document management and editing highlighting their main characteristics and downsides. We will continue in Section 2 with the theoretical approach of our system and then discuss in Section 3 the implementation and the technical details concerning it. Based on the model we’ve proposed, in Section 4 we shall present some conclusions and further work.

2 THEORETICAL APPROACHES

In this paper we shall present our implementation of such a system that aims to cover the aforementioned requirements. Dante is a web based system designed to be a good support for virtual teams in elaborating scientific papers. It offers document management facilities and process automation for repetitive tasks. Since all the data the system uses is stored in XML files, Dante can be document and user centered, allowing authors to easily edit, review and export in different formats their work. Teams are building around the document allowing them to be goal directed; all members of the team having the same rights. Each author is responsible of editing different chapters of the document, chapters on which the others could only place comments, content changing not being allowed, avoiding conflicts and stepping on each others and allowing individuals to reconcile with the teams goals. The application facilitates communication through synchronous and asynchronous channels. Documents being stored in XML files and using a web based user interface makes the system work on different platform and allows users to export documents in different open formats, representing a combination of best practices specific for the previous discussed systems. At this moment the application offers no version control capabilities, and therefore authors could not revert documents to older versions.

As mentioned earlier, Dante is both a user and document centric system, supporting collaboration in virtual teams and efficient document management. As described by Millward and Kyriakidou (2004), it is important for virtual teams to be a “singular concrete entity” with the following characteristics: stability, regular interaction, symbiosis and member proximity. Following this requirements, in Dante teams are organized around the document that they are creating. Member proximity results from the fact that each member can view the most recent version of the chapters that the others have created and that everybody knows who is responsible for a particular chapter. Each author can review others work and make suggestions related to each piece of text using the commenting tools. The symbiosis of the team is supported by the fact that each person’s responsibilities are clearly drawn and all members have the same rights, all depending upon others in improving their work. Interaction between members is supported by both synchronous and asynchronous mechanisms.

In collaborative real time editing systems several users can edit a file using different computers. An important aspect for this type of systems is group awareness (GA). GA provides users information about the status of a document and changes made by others. As shown in Raikundalia and Zhang (2004), several GA techniques have been identified: telepointers (multiple cursors are shown within the document), radar views, multi-user scrollbars and structured-based multi-page view, point jumping mechanism and user info list. In Dante, the GA problem is solved by using a structured-based multi-page view panel for displaying a project. One of the main downsides of all collaborative editing systems (CES) is that the document is saved on the document owner’s computer, all others participants being allowed to save a copy of the document leading in
time to lost edits and versioning problems. **Dante** being a web-based system has the advantage of delivering file storage facilities. All changes made by the authors are saved on the server which will offer only the most recent version of the document. Authors edit different parts of the document and send changes to the server via AJAX.

As shown in Leone, Hodel and Gall (2005), combining CES and DMS can result in greater performance improvements. **Dante** takes a similar approach providing solutions and process automation for creating, storing, retrieving, editing, and exporting documents collaborative, anytime on the most used platforms. All documents have metadata to easily manage them and provide extra search capabilities. Another important aspect is the presence on the internet and the collaboration with persons from outside the team using e-prints. Lawrence (2001) shows that articles freely available online are more highly cited and recommends, in order to achieve greater impact and faster scientific progress, that authors should aim to make research easy to access. Content can easily be transformed into HTML and allow others to post comments on the article if the team wishes so.

![Figure 1: Dante architecture.](image)

To summarize, the main characteristics that we consider essential for an efficient collaborative document creation and review system are:

- support for virtual team efficiency
- user and document centered
- support users location and time independency
- platform independent
- efficient group interaction
- allow users to concentrate on the content and waste as little as possible time with the technology that supports the process
- document management facilities

The presentation layer defines the following main tools: **Editor**, **Partners**, and **Chat**. The **Editor** allows users to manage old projects, start new ones, collaboratively edit current projects and export documents in different open formats and format the content using predefined templates. When starting a new project, the project owner must define a set of properties, like name (works like an identifier for the project), type (according to the type that has been selected, a certain template will be used to format the content when users export the document) and a list of members that will participate to the project (after defining the project, all participants will have the same rights) (Leone, Hodel and Gall 2005). During development extra sets of metadata will be added to the project allowing users to consult the state of the project, last modified date etc. Each author can edit one or more chapters from the document and place comments on those chapters edited by the others. Chapters are presented as elements in a tree menu and can be accessed in different windows. When accessing a particular chapter, the system checks the metadata associated to it and determines if the current user is the author of the chapter and displays a new menu that allows him to edit the content if so, otherwise allowing him only to add new comments to the component elements or edit comments define earlier. The editor is not made up from elements that will allow authors to format the content (i.e. defining font types, paragraph alignment etc.), but from elements that represent structural components of document (i.e. paragraph, note, quote, table etc.). The formatting of document will be done automatically by the application according to the template that the document owner had defined. Each structural component of the document can be commented by the rest of the team, the editor defining special zones at the end of each element where these comments can be consulted. When accessing chapters for whom the current users is not the owner, a different editor is loaded allowing the user to define and edit the content.
comments and reply to those that others have defined. Before finalizing the project, the chapters can be exported in different open formats (i.e. PDF) or to a link in order to be accessed by those that are not participating to the project.

The Chat sections allows user to communicate using synchronous channels. The section defines two main channels: Groups and Personal. The Groups channel lists all active projects for the current user and allows him to communicate with the members of each project in different panels. When a particular group is selected, the user can communicate with the members of that project that are online, the messages being available for all co-workers. All messages are stored in XML files and associated to the project, allowing user to consult the discussion archive or the project anytime. On the other hand, the Personal channel allows user to communicate with each friend defined in the Agenda using private channels. These messages are not stored in the message archive, only if the users decide so.

The Partners section has two main subcomponents: Agenda and Invitations. The Agenda allows users to manage their partners’ contacts and export them to micro-formats like HCard. The user’s personal data can be exported in VCard format. The Invitations subcomponent manages the user’s invitations to participate at different projects (only after accepting an invitation a user can actively participate to a project). All personal data is stored in XML files, allowing the system to easily integrate them into the projects content when exporting a project in a final state according to different templates.

In the next section we shall discuss the details regarding the implementation of our web based collaborative document creation and review system.

3 IMPLEMENTATION

As previously mentioned, Dante is a web based system design upon de model-view-controller design-pattern which stores all data in XML files. The View layer personalizes the user interface according to user’s rights and permissions and displays the appropriate editor after reading the documents metadata. It also implements a communication module which transfers data to the Controller layer using AJAX. The Controller layer handles the events triggered from the UI and calls the appropriate handler from the Model layer. We will concentrate our attention on the Model layer which manages all XML documents.

The Model layer consists mainly from two subcomponents: the Document manager and the Repository. The Repository is a collection of DTD, XML and XSLT files used by the Document manager module.

The Model layer consists mainly from two subcomponents: the Document manager and the Repository. The Repository is a collection of DTD, XML and XSLT files used by the Document manager module.

![Figure 2: Document workflow.](image)

The document type definition files describe the structure of the documents and all metadata that can be added to them. In this case, to describe the structure of the scientific papers we’ve created a document type called xmlDocument. The structure of this type is in a way similar to the more mature DocBook xml vocabulary (Walsh 2005), being more simplified and particularized for scientific papers. The xmlDocument DTD defines the structure of the whole document viewed as a project. The two main elements of an xmlDocument are docInfo, which describes the metadata associate with the document (i.e. document name, authors, document type, last update, version etc.) and docBody. The docBody defines the content of the document (i.e. abstract, references, appendices etc.) and for each chapter the metadata associated (id, title and author) and a link to the XML file that defines its content. This separation of each chapter in different XML files allows a more flexible management of content and metadata. The vocabulary that describes the chapter’s elements is defined by the xmlChapter document type definition, implementing means of storing and identifying elements. As mentioned earlier, each author can define comments for the chapters that he is not editing. These comments are stored in separate XML files, and the vocabulary describing them defines elements to uniquely identify de comment, describe its characteristics (i.e. author, date etc.) and link it to the element to which it referees.
In order for the Document manager to easy set the corresponding roles and rights, a document type definition is implemented to store the projects. This DTD allows defining for each user the projects for which he is a collaborator or the owner. The Document manager module consist in several classes based on the DOM implementation which automates the processes concerning the creation, storage, retrieval and the export of the documents. All server side scripts have been implemented in PHP which includes a good support for XML, complying with the commonly used standards (SAX, DOM, SimpleXML, XMLReader, XMLWriter, and the XSLT). Some of the classes defined in the document manager come as a wrapper for the DTD residing in the repository: the xmlChapter class, for example, handles the creation and the update of different document objects, prepares the content for printing accessing the appropriate class when exporting (i.e. the pdfPrinter class when exporting to PDF). This module generates the appropriate user interface according to the user’s role and rights, loading the corresponding XSL and parsing the requested XML file. For the same DTD the Repository defines different XSL transformation files for each role and user-level, restricting the access to different actions that the user can perform upon the document. Most of the UI used by Dante is generated using XSL transformations.

The Chat module works in a similar way, defining a DTD for storage purposes and different XSL transformation schemes for display. As mentioned previously, the user can communicate using private channels or rooms dedicated to different projects. The messages exchanged between users in these project related rooms are bound to the project using metadata (although physically residing in different locations) allowing project based message archiving. This approach allows members to catch-up with the team when not being able to join the group. The Agenda module stores member’s personal data (this being done also using XML files) and allows users to export this information in HCard and VCard microformats. All templates require a minimum of identification data for the authors, the project importing all required data when formatting from this module. We have tried to structure the implementation as much as possible according to functionality of the whole system, defining a modular structure.

The user interaction is managed using JavaScript. Because the communication from the server to the client is mostly done using XML chunks (or an entire file), the classes defined on the client-side are in some manner a transposing of those residing on the server being also a wrapper upon the document. The Session Manager module from the Controller layer determines which script to be loaded according to conjuncture determining the operations that a user can execute on a particular document. When editing a particular chapter, all new elements and all changes are stored in different queues and only when the users decides to save the document the content of these queues is send using AJAX to the server. These classes are also based on the DOM model, the content being sent to the server representing a DOM node to be inserted in the document residing on the server. The queues are gradually discharged when a component element is successfully saved on the server. This approach allows defining specific UI elements and behavior to each document type being treated. The user screen is blocked while saving the content of the queues, ousting the chances to make unsafe changes to the content. When appropriate, an XSLT processor is used on the client-side to reduce the charge on the server.

We have chosen this modular approach based on document types and user interfaces particularized on roles/rights and document types in order to facilitate further development of the system in such a manner that it could handle a lot more document types.

4 CONCLUSIONS

The work reported in this paper has addressed the problem of integrating document management functionalities and workflow capabilities into single system with support for virtual teams in order to achieve a very efficient solution for collaborative document editing. Allowing users to edit collaboratively documents and supporting their needs for time and location independence can result in more team efficiency. Users have to be able to concentrate their efforts on content and reduce as much as possible the time used to handle and integrate technologies. An approach that integrates functionalities from document management systems, workflow management and collaborative editing proves to be a much more user centered and supportive collaborative solution. As such an implementation, Dante tries to offer sufficient support in order for the teams to be goal directed and efficient. Using a web based solution does not impose restrictions on the users regarding operating systems or software. Taking in consideration the multitude of solutions available, we have to consider
using as much as possible open formats in order for the content to be easily integrated. As shown in this paper, using XML as building blocks allows the system to be document and user centered. We have presented a case study of a particular field where such an implementation would be a great support. As further work, we intend to develop our system toward a framework that can handle a wide range of documents, taking in consideration the increased need for collaborative document editing in a multitude of working fields.

ACKNOWLEDGEMENTS

The work presented has been funded by the research grant “Intelligent System for Business Decisions Support”, Director Nitchi Stefan, PhD., Professor, PNII Program, 91-049/18.09.2007 supported by Higher Education Ministry.

REFERENCES


ANTECEDENCES AND CONSEQUENCES OF E-SERVICE QUALITY ACROSS INDUSTRY SECTORS

Dauw-Song Zhu
Business Administration & Accounting Department, National Dong Hwa University
No. 1, Sec. 2, Da Hsueh Rd. Shou-Feng Hualien, Taiwan 974
dswu@mail.ndhu.edu.tw

Chih-Te Lin
Business Administration Department, National Dong Hwa University & Lecturer of Food and Beverage Management
Department, Taiwan Hospitality & Tourism College
No. 2, Lane 46, Jhongmei 3rd St, Hualien, Taiwan 970
thomsonhua@gmail.com

Yu-Ling Su
Business Administration Department, National Dong Hwa University & Lecturer of Food and Beverage Management
Department, Taiwan Hospitality & Tourism College
No. 2, Lane 46, Jhongmei 3rd St, Hualien, Taiwan 970
yuling@tht.edu.tw

Keywords: e-Service quality, e-commerce quality, online quality, Internet quality.

Abstract: A few studies are available on the different sectors basis on the determinants of success in online retailing. This study tries to identify which e-Service dimensions serve as determinants of e-Service quality (Zeithaml et al., 2002). Therefore we focused on analyzing impacts on overall e-Service performance of various e-Service quality dimensions in internet retailing. We analyzed the impact of e-Service quality dimensions of online retailers along several sectors. We found that while many dimensions of online service performance were similar in their impact across all sectors, several dimensions in particular exhibited sector-by-sector performance differences.

1 INTRODUCTION

What companies on the Internet require is to deliver high value to the customer, build customer loyalty, encourage repeat purchases and maintain long term relationships with the customers (Gurau 2003; Parasuraman and Zinkham 2002), since recent research indicates that ‘positively predisposed on-line customers’ demand much more (Van Riel et al. 2003). To do so, as Zeithaml et al., (2002, p362) argue, companies need to shift the focus of e-business from e-commerce to e-Service. Such shift implies understanding what customers’ value in an on-line encounter, how they perceive the experience and how they evaluate its quality (Lauren et al, 2006).

Understanding, measuring and managing e-Service quality on the web has thus become an issue for ensuring customer satisfaction, loyalty and firms’ profitability and service, both for industry and the public sector (Buckley 2003). Thus, we examine each of several industry sectors separately to distinguish the web site dimensions that are associated with creating overall satisfaction in the online consumer. Here we concisely review literatures that have empirically examined antecedents and consequences dimensions of e-Service quality in electronic services. Table 1 is a list of the variables that have been examined in these literatures, both in multi-industry sector e-retailing literatures, and in single industry studies related to general e-Service, online banking/ financial services, travel services, book retailing, and health care services.
2 ANTECEDENCES

A few previous literatures have studied the issue of online e-Service quality from a multi-industry perspective, or have collected data across several industry sectors. According to review of 19 literatures, we discovered overall web site quality is the most frequent antecedent across general e-Service/ e-retail industry (Kim et al., 2002; Palmer, 2002), online banking/ financial services industry (Chen and Hitt, 2002; Montoya-Weiss et al., 2003; Jayawardhena, 2004), travel industry (Harris and Goode, 2004), books industry (Gefen, 2002) and health industry (Van Riel et al., 2001). The second frequent antecedent is trustworthiness or assurance across four industries of online banking services industry (Balasubramanian et al., 2003; Jayawardhena, 2004), travel industry (Harris and Goode, 2004), books industry (Gefen, 2002; Lee, G. G. & Lin H. F., 2005; Lin, H. F., 2007) and health industry (Viitanen et al., 2003).

We examined general e-Service industry, online banking services industry, travel industry and books industry, and found that security of environmental risk is a antecedent variable of general e-Service/ e-retail industry (Kim et al., 2002; Zhang, X. & Prybutok, V. R., 2005), online banking industry (Balasubramanian et al., 2003; Montoya-Weiss et al., 2003), books industry (Gefen, 2002) and in health industry (Viitanen et al., 2003). In addition to that, we also found content or information is the predicator of e-Service quality across industries of e-retail industry (Kim et al., 2002; Palmer, 2002; Janda, S., Trocchia, P. J. and Gwinner, K. P., 2002; Liao, C., Palvia, P. and Lin H. N., 2006), online banking industry (Montoya-Weiss et al., 2003) and books industry (Lin, H. F., 2007). Furthermore, personalization, customization or interactivity were predicator across e-retail industry (Palmer, 2002), online banking industry (Chen and Hitt, 2002; Jayawardhena, 2004), travel industry (Van Riel et al., 2003) and books industry (Lin, H. F., 2007). But technical quality or technical adequacy viewed as antecedent only in e-retail industry (Choi et al., 2004; Kim et al., 2002; Liao, C., Palvia, P. and Lin H. N., 2006) and health industry (Viitanen et al., 2003).

We found accessibility or e-Service convenience used as independent variable in studies of e-retail industry (Zhang, X. & Prybutok, V. R., 2005), online banking industry (Jayawardhena, 2004) and travel industry (Van Riel et al., 2003). Fulfillment or reliability is the antecedent across travel industry (Van Riel et al., 2003) and books industry (Gefen, 2002; Lee, G. G. & Lin H. F., 2005), functional quality is the antecedent across e-retail industry (Choi et al., 2004; Kim et al., 2002) and health industry (Van Riel et al., 2001), and security or privacy is the antecedent across e-retail industry (Kim et al., 2002; Janda, S., Trocchia, P. J. and Gwinner, K. P., 2002) and books industry (Lin, H. F., 2007).

For banking and travel industry, aesthetic appeal is the driver of e-Service quality (Montoya-Weiss et al., 2003; Van Riel et al., 2003), and comparative prices is also the driver across e-retail industry (Choi et al., 2004) and banking industry (Balasubramanian et al., 2003). Download speed became the antecedences across e-retail industry (Palmer, 2002) and travel industry (Van Riel et al., 2003). Usefulness is significant driver in books industry (Koufaris, 2002), health industry (Viitanen et al., 2003). And on time delivery became the significant driver in books industry (Gefen, 2002) and health industry (Viitanen et al., 2003). Online banking industry and travel industry studies used navigation as antecedent (Montoya-Weiss et al., 2003; Van Riel et al., 2003). Product quality viewed as predicator in e-retail industry (Choi et al., 2004) and travel industry (Clemons et al., 2002), responsiveness used as predicator in books industry (Lee, G. G. & Lin H. F., 2005; Lin, H. F., 2007). However, web site design is significant association with dependent variables only in books industry (Lee, G. G. & Lin H. F., 2005; Lin, H. F. 2007).

3 CONSEQUENCES

We found satisfaction is the most major consequent variables in the previous studies across general e-Service/ e-retail industry (Kim et al, 2002; Palmer, 2002; Janda, S., Trocchia, P. J. and Gwinner, K. P., 2002; Zhang, X. & Prybutok, V. R., 2005), online banking/ financial services industry (Balasubramanian et al., 2003; Montoya-Weiss et al., 2003), travel industry (Van Riel et al., 2003; Harris and Goode, 2004), books industry (Lee, G. G. & Lin H. F., 2005; Lin, H. F., 2007) and health industry (Van Riel et al., 2001; Viitanen et al., 2003). And the second major consequences is customer loyalty across four industries including general e-Service/ e-retail industry (Kim et al., 2002; Palmer, 2002), travel industry (Van Riel et al., 2003; Harris and Goode, 2004), books industry (Gefen, 2002; Koufaris, 2002) and health industry (Van Riel et al., 2001; Viitanen et al., 2003). Next consequent variable is quality or e-Service quality studied in...
three industries including online banking/financial services industry (Jayawardhena, 2004), travel industry (Van Riel et al., 2003) and books industry (Lee, G. G. & Lin H. F., 2005). Moreover, value variable is studied across different industries general e-Service/ e-retail industry (Choi et al., 2004), travel industry (Harris and Goode, 2004) and health industry (Van Riel et al., 2001).

In addition, we examined online banking and books industry, and found customer switching is the antecedent separately studied in Chen and Hitt (2002) and Gefen (2002). And Clemons et al. (2002) surveyed travel industry, price dispersion as antecedent. Otherwise, Janda, S., Trocchia, P. J. and Gwinner, K. P., (2002) examined general e-Service/ e-retail industry and used likelihood of future purchase, likelihood of complaining and word-of-mouth as dependent variables, but Liao, C., Palvia, P. and Lin H. N. (2006) used trust as dependent variables at the same industry. We may make up the situation of outcome variables were identified, the summary shows in Table IV.

4 METHOD

The a-priori theory underlying our analysis of the relationship between individual e-Service quality dimensions and overall e-Service performance is that customers in each e-retailing sector will possess certain preferences about aspects of an online shopping system, and e-Service designers will strive to understand those sector-specific factors and will design their online service system to attempt to satisfy customers in their sector. Across all online retailing sectors, a large body of e-Service quality dimensions will exist, subsets of which will be relevant for managing individual sectors.

5 CONCLUSIONS

Firms today need to understand better how different industry sectors and contexts lead to different relationships between tactical activities such as e-Service quality management and their overall performance outcomes (Bolton et al., 2004). With the phenomenal growth of e-Services, research must aim to better understand the dimensions, antecedences, and consequences of quality and customer satisfaction in the online domain across various industry sectors (Lauren et al, 2006). Therefore we focused on analyzing impacts on overall e-Service performance of various e-Service quality dimensions in internet retailing. We analyzed the impact of e-Service quality dimensions of online retailers along several sectors. We found that while many dimensions of online service performance were similar in their impact across all sectors, several dimensions in particular exhibited sector-by-sector performance differences.

This study makes several contributions to the literature on e-Service performance specifically due to our sector-by-sector analysis. In prior studies, researchers have taken a scattered approach to identifying e-Service quality dimensions for their studies. They have also focused on the issue from a single industry perspective. This research approach has led the literature to a state of many quality dimensions having been identified, but little evidence that can be triangulated across sectors or across studies. In closing, we hope our paper will prove insightful to academics and managers interested in managing e-Service operations. We believe there is much interesting research left to be done on such topics, and hope our paper serves to motivate others to explore the emerging issues in this interesting industry arena.

REFERENCES


RFID PASSWORD MANAGEMENT METHODS FOR FALSIFICATION PREVENTION IN BOOKSTORE MANAGEMENT USING SECURE RFID TAGS

Yuichi Kobayashi, Yoji Taniguchi
Hitachi, Ltd., Systems Development Laboratory, 1099 Ohzenji, Asao, Kawasaki, Kanagawa, 215-0013, Japan
yuichi.kobayashi.ma@hitachi.com, yoji.taniguchi.rf@hitachi.com

Toshiyuki Kuwana
Hitachi, Ltd., Tracing & Tracking Systems Division, 890 Kashimada, Saiwai, Kawasaki, Kanagawa, 212-8567, Japan
toshiyuki.kuwana.jr@hitachi.com

Masanori Akiyoshi
Graduate School of Information Science and Technology, Osaka University, 2-1 Yamada-oka, Suita, Osaka, 565-0871, Japan
akiyoshi@ist.osaka-u.ac.jp

Keywords: RFID, falsification prevention, risk analysis.

Abstract: The receipt data on radio frequency identification (RFID) tags attached to books may be used to prevent shoplifting in bookstores. To protect the receipt data, it is important to manage the passwords of RFID tags. We use Secure RFID tags, which protects data with an RFID password, for preventing falsification of RFID data. We also propose ten methods that manage RFID passwords from the point of the phases of RFID passwords and the ways a password is associated with a Secure RFID tag. We analyze and compare these methods using fault tree analysis. We show that our proposed RFID passwords management methods are effective in preventing falsification for resale.

1 INTRODUCTION

The shoplifting of books is a serious problem in Japan. A report says that there are many shoplifted books for resale. The publishing industry in Japan is examining a method for attaching an RFID tag to books, which records receipt data on the RFID tag memory. However, there is a risk of someone reselling a shoplifted book to a second-hand bookstore, after illegally overwriting the receipt data in the RFID tag. So a method for protecting the data in the RFID tag is needed.

There are several methods that mount hash logic on an RFID tag and authenticate the user to protect the RFID data (Weis 2003; Engberg, Harning & Jensen 2004; Tripathy & Nandi 2006). However, these methods are not realistic in a situation in which an RFID tag is attached to a book because mounting advanced calculation logic on an RFID tag requires a larger IC chip, which increases the cost of the RFID tag.

Secure RFID tags are low cost RFIDs. This type of RFID tag protects data using a simple password authentication method. We chose to use Secure RFID tags because these tags are smaller, cost less, and suitable for attaching to books. It is very important to manage an RFID password when using Secure RFID tags. Therefore we propose ten methods for managing RFID passwords. We analyze their security, and show that they are effective in preventing falsification.

2 BOOK MANAGEMENT SYSTEM WITH SECURE RFID TAGS

In this chapter, we describe the book management system with Secure RFID tags in a bookstore environment.
2.1 Book Distribution Procedure for Preventing Shoplifting

The publishing industry in Japan is considering a method for preventing shoplifting by using Secure RFID tags. Figure 1 shows the procedure for preventing shoplifting in a book distribution. First, a bookstore writes in “Stock” onto the Secure RFID tag when a book arrives from a distributor. This denotes the state before being sold. Next, when a consumer buys the book, the bookstore changes “Stock” to “Sold” in the Secure RFID tag. At the entrance of the store, the security system reads the data on the Secure RFID tags to check if the book was purchased. Finally, when a second-hand bookstore buys a book from a consumer, it checks whether the book’s Secure RFID tag reads “Sold” or “Stock”. Not only bookstores, but second-hand bookstores cooperate with the publishing industry. All the companies relevant to book publishing and selling can examine the system to counter the shoplifting problem.

In order for this system to work, the data of the Secure RFID tag must not be able to be easily rewritten from “Sold” to “Stock”. Rewriting information should be done with an access control function with a password of the Secure RFID tag.

2.2 An Existing RFID Passwords Management Method

Figure 2 shows the RFID password management system for a bookstore.

Operation 1: When the book arrives, the book’s code written in the Secure RFID tag is read, and the pre-arrival data is checked. After it is checked, the book code is input into the server with the arrival data. Then, the date of arrival, bookstore code, “Stock”, etc... are written onto the Secure RFID tag. An RFID password is then set up for the Secure RFID tag, and the memory is locked using that password.

Operation 2: When the book is sold, the book code in the Secure RFID is read, the price is checked, and the sale information is input into the server. At this time, the memory of the Secure RFID tag is unlocked by using the RFID password, and “Sold” is rewritten onto the Secure RFID tag. Finally, the memory is locked again by using the password.
Operation 3: A gate type interrogator system at the entrance is set up, and the interrogator checks if “Sold” or “Stock” is written on the Secure RFID tag.

Operation 4: If the book is returned, the memory of the Secure RFID tag is unlocked by using the RFID password, and all data from the bookstore and the password are deleted. Afterwards, the returned book’s information is input into the server.

The publishing industry in Japan tries to manage RFID passwords by using identical passwords in all bookstores, and applies the RFID password management method for each bookstore. This method has the advantage of being less expensive; however, if one RFID password is stolen, all Secure RFID tags are compromised. The possibility of information being falsified on the Secure RFID tags attached to books stocked in all bookstores is high. Therefore, the distinction between a shoplifted book and a legally bought one becomes difficult in a second-hand bookstore, and the effect of preventing theft by using the Secure RFID tag weakens. It is necessary to manage RFID passwords more carefully.

3 RFID PASSWORDS MANAGEMENT METHODS

We propose ten methods for managing not identical RFID passwords but different passwords so that the risk after a password is stolen will be lessened. It is effective to divide RFID passwords into groups to manage different RFID passwords. For example, the group includes an inspection interrogator, a book, etc... Moreover, because it is usually impossible to read the RFID password set in the Secure RFID tag, it is necessary to set the RFID tag with an RFID password beforehand. Therefore, we present ten methods which can be used to manage RFID passwords from the point of a phase of RFID password management and the ways in which a password is associated with a Secure RFID tag. Figure 3 shows the relationship between the phase of RFID passwords, and the ways passwords are used for a Secure RFID. There are four phases of RFID password management, (1) each bookstore, (2) each inspection interrogator system, (3) each arrival, and (4) each book. On the other hand, there are (a) the way of using a conversion table, (b) the way of using operation logic, and (c) the way of updating a key.

We define the combination of phases of RFID passwords and the ways a password is used for a Secure RFID tag as the RFID password management methods.

3.1 Phases for Managing RFID Passwords

A bookstore that manages RFID passwords separately by each group lessens the risk of RFID tag falsification more than a bookstore that manages identical RFID passwords. We explain the phases for managing RFID passwords as follows.
(1) Each bookshop: Manage the different RFID passwords in each bookstore. Each bookstore decides and manages an RFID password.

(2) Each interrogator system: Manage the different RFID passwords in each interrogator system in a bookstore when the book arrives. Each interrogator system sets up an RFID password and an interrogator’s ID to the Secure RFID tag attached to the newly arrived book.

(3) Each interrogator system and arrival: Manage the different RFID passwords in each interrogator system in a bookstore when the book arrives and update the RFID passwords each time books arrive. Each interrogator system sets up an RFID password, an interrogator’s ID, and the arrival date to the Secure RFID tag.

(4) Each book: Manage the different RFID passwords in each Secure RFID tag attached to the book. The bookstore associates the RFID password with a unique ID that identifies the Secure RFID tag.

It is possible to manage the different RFID passwords in each title of the book. However, the titles with only one book in stock accounts for 70% of the inventory in most bookstores. Therefore, we omit the explanation for each title here because each title and each book are almost the same.

3.2 Ways in which RFID Passwords are Associated with a Secure RFID Tag

When a different RFID password is set to a Secure RFID tag, the system should associate that password with that Secure RFID tag. We explain the ways in which this is accomplished.

(a) Conversion table: The system generates the RFID password at random, and associates the RFID password with the interrogator’s ID, arrival date, or a unique ID written to the Secure RFID tag to identify the phase of RFID password management.

(b) Operation logic: Use operation logic with an electronic key. The system generates an RFID password by calculating data, such as the interrogator’s ID, arrival date, or a unique ID written to the Secure RFID tag (Kobayashi, Kuwana, Taniguchi & Komoda 2007).

(c) Update electronic key: Periodically update the electronic key used in (b).

4 SECURITY EVALUATION OF RFID PASSWORD MANAGEMENT METHODS

In this chapter, we show the results form a fault tree analysis (FTA) of the RFID password management methods, and compare those analysis results.
4.1 Fault Tree Analysis of RFID Passwords Management Methods

We analyzed the security performance of the RFID password management method, which combined the phase of an RFID password and the ways in which the password is associated with a Secure RFID tag described above, using FTA.

We know that the probability of a successful falsification of RFID data for resale is low; the RFID password management method will have a high security performance. Therefore, let a top event of FTA be the probability that the RFID data of a shoplifted book will be falsified, and resale of the book will be successful in one month. Moreover, it is necessary to decipher the RFID password of the Secure RFID tag so that someone may falsify the RFID data. There are two effective attacks for deciphering an RFID password. The first is that someone illegally uses a common interrogator and hacks into the Secure RFID tag like brute force attack. The second attack is that someone steals and illegally uses the handy interrogator of the bookstore. The FTA is shown in Figure 4. The probability of each event of FTA is explained below.

(B-11): This event is the probability of falsifying the RFID data of a bookstore using the RFID password stolen at another bookstore. In the existing RFID password management method, this probability $P_{B-11}$ is 100% because an RFID password is identical at all the bookstores. In the proposed RFID password management method, this probability $P_{B-11}$ is 0% because the RFID passwords at least differ for each bookstore.

(B-12): This event is the probability that an RFID password will be decoded in one month by hacking into a Secure RFID tag. This probability $P_{B-12}$ was obtained from the following expression.

$$P_{B-12} = \frac{1}{E(Time)} = 4.1\%$$  \hspace{1cm} (1)

$E(Time)$ is an expected time spent on the hacking attack. The time that a Secure RFID tag is checked whether one password is right using one interrogator is assumed to be about 30 milliseconds and the length of the Secure RFID password is assumed to be 32 bits according to the specification of Secure RFID.

(B-2): This event is the probability of falsifying sales data on a Secure RFID tag in one month using a common interrogator. This probability depends on the number of books that can be falsified using a decoded RFID password. This probability $P_{B-2}$ is 100% in the case of the existing method or “(1) each bookstore” because all RFID passwords in the bookstore are the same. In the case of “(2) each interrogator”, this probability $P_{B-2}$ is 33% when there are three interrogators for inspection is three. In the case of “(3) each interrogator and arrival”, this probability $P_{B-2}$ was obtained by the following expression.

$$P_{B-2} = \frac{1}{IN} \times \frac{ABN}{BN} = 27\% \hspace{1cm} (2)$$

where $IN$ denotes the number of the interrogators for inspection, $ABN$ denotes the number of applicable stocked books, and $BN$ denotes the number of stocked books. The applicable stocked book means the books that remain unsold for one month because the
decoded RFID password is associated with those books. IN is assumed to be three and ABN/BN is assumed to be 0.79 because the merchandise turnover is about 17%, and the update frequency of the RFID passwords is once a week. In the case of “(4) each book”, this probability $P_{B-2}$ is 1% when the number of books stolen in one month in one store is 100, because the number of books that can be falsified by using the decoded RFID password is only one.

(C-1): This event is the probability that an interrogator will be stolen from a bookstore in one month. Here, this probability $P_{C-1}$ is assumed to be 0.3%.

(C-2): This event is the probability of falsifying sales data on a Secure RFID tag in one month using the stolen interrogator. This probability depends on the ways in which an RFID password is associated with a Secure RFID tag. In the case of “(a) conversion table”, this probability depends on the managing phase of RFID passwords because the conversion table is contained in the interrogator of the bookstore. This probability $P_{C-2}$ is 100% in the case of “(1) each bookstore”, 33% in the case of “(2) each interrogator” or 27% in the case of “(3) each interrogator and arrival”, as well as in the case of the (B-2) event. In the case of “(4) each book” this probability $P_{C-2}$ is 76% because the merchandise turnover is about 17% and the update frequency of the RFID passwords is once a day. In the case of “(b) operation logic”, this probability $P_{C-2}$ is 100% because the operation logic is contained in the interrogator of the bookstore. In the case of “(c) update electronic key”, this probability can be estimated at 80% when changing the key once a week.

The result of FTA for each RFID password management method is shown in Table 1. The RFID password management methods below from “(1) each bookstore” is classified into the phases of RFID password management and the ways in which a password is associated with a Secure RFID tag. The falsification risk in Table 1 is the probability that falsification of the RFID data for resale will be successful. This falsification risk is low because the security of the method is high.

### 4.2 Comments of the Results of FTA

Figure 5 shows the probability that a falsification will be successful for each RFID password management method. This figure also shows that the difference between the phases of RFID password management is larger than the difference between the ways in which a password is associated with a Secure RFID tag. We found that it is better to manage RFID passwords in several phase. Moreover,
all the RFID password management methods are effective in preventing falsification for resale because the highest probability that falsification will be successful for “(1) each bookstore” is 4.25% for all the methods.

Table 1: Comparison of the security risk for each RFID password management method.

<table>
<thead>
<tr>
<th>RFID passwords management method</th>
<th>Falsification Risk</th>
</tr>
</thead>
<tbody>
<tr>
<td>Identical</td>
<td>100.00 %</td>
</tr>
<tr>
<td>(1) Each bookstore</td>
<td>4.25 %</td>
</tr>
<tr>
<td>(a) Conversion table</td>
<td></td>
</tr>
<tr>
<td>(2) Each interrogator</td>
<td>1.42 %</td>
</tr>
<tr>
<td>(3) Each interrogator &amp; arrival</td>
<td>1.13 %</td>
</tr>
<tr>
<td>(4) Each book</td>
<td>0.24 %</td>
</tr>
<tr>
<td>(b) Operation logic</td>
<td></td>
</tr>
<tr>
<td>(2) Each interrogator</td>
<td>1.58 %</td>
</tr>
<tr>
<td>(3) Each interrogator &amp; arrival</td>
<td>1.31 %</td>
</tr>
<tr>
<td>(4) Each book</td>
<td>0.30 %</td>
</tr>
<tr>
<td>(c) Update electronic key</td>
<td></td>
</tr>
<tr>
<td>(2) Each interrogator</td>
<td>1.53 %</td>
</tr>
<tr>
<td>(3) Each interrogator &amp; arrival</td>
<td>1.26 %</td>
</tr>
<tr>
<td>(4) Each book</td>
<td>0.25 %</td>
</tr>
</tbody>
</table>

5 CONCLUSIONS

We proposed methods that use Secure RFID tags and RFID password management for preventing falsification of RFID data in book distribution. These proposed methods were explained by the point of phases of RFID passwords, and the ways in which a password is associated with a Secure RFID tag. We showed that these methods decreased the probability of falsification to about 5% or less in this situation. These results are useful for RFID password management in bookstores.

ACKNOWLEDGEMENTS

This paper is based on the achievement of a Japanese National Research and Development Project, the “Secure RFID Project” that was conducted by METI (Ministry of Economy, Trade, and Industry) for the eight months from August 2006 to March 2007.

REFERENCES


WORK LISTS FOR THE TRANSPORT OF PATIENTS

A Case for Mobile Applications in Health Care

Andreas Holzinger, Jürgen Trauner
Institute of Medical Informatics, Statistics and Documentation (IMI), Research Unit HCI4MED
Graz University Hospital, Auenbruggerplatz 2/V, A-8036 Graz, Austria
andreas.holzinger@meduni-graz.at

Stefan Biffl
Complex Systems Engineering Lab, Institute of Software Technology and Interactive Systems (IFS)
Vienna University of Technology, Favoritenstrasse 11, A-1040 Wien, Austria
stefan.biffl@tuwien.ac.at

Keywords: Mobile User Interfaces, Medical Workflow optimization, Mobile Computing in Hospitals.

Abstract: In many hospitals, the workflow involved in transporting patients is supported by interactive work lists on desktop computers that provide the status of a work order and detailed information on request. However, the nurses, who organize and conduct the transportation, need to go to the desktop in order to update the work list status; which has been found very time consuming. In this paper we report on the development of a mobile solution prototype that provides interactive transport work lists on a PDA and we discuss some design and implementation issues. Despite the limited screen size, end-users were able to use the PDA with no more problems than their usual desktop user interface as the PDA user interface can be customized with property files to fit exactly the requirements of the end-users in the hospital.

1 INTRODUCTION

The transport of patients within hospitals is often supported by work lists on desktop computers (Gale & Gale, 2000). The central information is the status of a work order. Whilst such a non-mobile system is useful for work orders that do not change, the highly dynamic nature of hospital work often results to changing situations (Holzinger & Errath, 2007). Consequently, the nurses, who organize such transportations, are required to go to the desktop computer in order to update the work list status – only to find out that their next patient is waiting where they just came from. In this context, a desktop solution is time consuming, and consequently wastes personal resources. In our project example, at the Danube Hospital Vienna, nurses have the core responsibility of transporting patients between the Radiology department and other medical departments: 14 different rooms, each with a separate list of appointments. At the beginning of their shift, the nurse prints out the list of patients to be transported on that day. Approximately 12 nurses are responsible for transportation per shift, which means that some must work with two lists. The list contains the name and the date of birth of the patient, for identification; the station where the patient lies; which type and when the checkup is planned; and whether the patient has to lie in bed or can walk. Usually, the nurse does not transport the patients in the order given on the list. Each transport is arranged with the assistant at the medical unit. Some changes in order are made by the nurse, for example, when more than one patient is at the same station or waiting at the station where another patient was returned. Then they may be fetched in order to save time. However, nurses are not automatically notified when patients are added during their shift, so either the nurses are informed about the new additions by the assistant or, less often, by telephone.

2 MATERIAL AND METHODS

In order to ensure platform independency of our mobile prototype, our first choice was Java. Fortunately, there is a free Virtual Machine for every Windows Mobile 2003 device (and later versions), called Mysaifu JVM (Freebeans, 2008).
Although this is not a Java 2 Micro Edition, it is a Java 2 Standard Edition (J2SE) conformant Virtual Machine, which is built as a Source Forge project under the GNU Public License Version 2 (GPLv2). For this project we used version 0.2. This version supports both the Abstract Windowing Toolkit (AWT) and Swing. We applied small sample programs in order to ensure full functionality; the use of a standard edition JVM has, in our opinion, some benefits: first, it is very simple to test the application before copying the files to the PDA. Second, J2ME is not just J2ME; which means that there are many different versions highly dependent on the performance of the device (Sun, 2008), (Knyziak & Winiecki, 2005). Because of these issues and of other restrictions of J2ME, the J2SE was the best way to implement this application, in order to provide compatibility with most of the current systems – which are usually developed in Java (Bruno, 2005). The general idea was to use an already available connection to the existing server. The main connection type is the Bright Side Framework (BSF) (Brightsidefactory, 2008), which transports XML content over the HTTP protocol with dynamic instantiation and invocation of available interfaces. However, in such specialized circumstances the JVM for the PDA is not fully compatible with the version for the PCs. The other connection was based on Remote Method Invocation (RMI), facing the problem that dynamic interfaces were not supported on the PDA. This required us to develop another connection for the PDA. Due to the communication protocol should not be changed in the first version we built a relay application.

3 BACKEND LESSONS LEARNED

A new concept was necessary, employing a relay, in order to establish the connection to the server. On the server side the connection was built using RMI or BSF (Sims, 2004), while, on the PDA side the communication is based on an object stream. The BSF and the RMI connections were given the means to access the data on the server. Both communication connections had to be adapted to fit the requirements of this application. The communication is no longer made over a static connection; this is due to the fact that the relay has to be able to open many connections. On top of this, some additional functions were necessary. The most interesting change is the function which returns the sessionId. The communication process between the relay and server, and relay and PDA client, is provided by SASTransferObjects. These made it necessary to add a function to transport them without alteration. Example: the logon data is stored within an object but the function takes the username and the password as its arguments, therefore it is necessary to prevent direct access to this data.

455

On the other hand, there is the client. The newly built communication method is also sub-classed from the SASServerProxy. This makes it possible to change the method of communication just by changing the properties file. The first idea was to use an object stream for both directions; however there was some difficulty with this method. Generally, the communication with objects from the client to the relay does not work – because it is far more complicated as it normally is on PCs, due to the different JVMs. As already mentioned, the Mysaiifu JVM is not yet fully compatible with the Java defaults, since it calculates the version numbers and does not allow for these to be fixed, as implemented in the source code. Every class which has a fixed version number causes an error at the relay when it is de-serialized. The InvalidClassException, states that the serial versions are different but in reality they are the same and only one number has been wrongly calculated. The solution for this is to change the version number before de-serialization. Any data which is sent through a stream is sent as a series of bytes. In an object stream, this is done with a special protocol which is publicly available (Sun, 2004). There is an exact description of how the class and the serialVersionUID are printed out. Some general information is followed by the class and super class information, which is succeeded by the class name printed in UTF. Directly after the name, the version number is written to the next eight bytes, followed by some flags and the variables, which of course can also be classes.

Once the object serialization stream protocol is known, this ceases to be a problem, since the right version numbers are known by the relay.

After the serialVersionUID of any class defined in the properties file is changed, the objects from the received byte stream can be regenerated and it is then possible to send a SASTranferObject, including classes, from the PDA client to the relay. However, in the other direction, the communication from the relay back to the PDA client is less simple.

The reason for this is mainly that the JVM on the PDA throws an InvalidClassException for each class. Theoretically, it would also be possible to change the version numbers on all the classes transferred on the PDA, practically, there are too many and the cost would be too high; therefore it was decided to send strings as a series of bytes. A
very simple protocol was invented where the string is divided into fields with field terminators. The PDA splits up the field and rebuilds the resulting set. This method has proved successful. The disadvantage for the relay, was the necessity of rethinking the data encryption technique, however, this actually turned out to be an advantage. Since the communication is built up with such easy methods, the necessary number of libraries decreased and now requires less than 100 KB.

The WPA encryption, which is supported by the WLAN hardware, is generally sufficient to secure the transmitted data. However, since the client on the PDA is used within a hospital and includes the transmission of sensible patient data (Weippl, Holzinger & Tjoa, 2006), (Miller, 2004), (Mupparapu & Arora, 2004) we decided to include additional encryption. The first technique we considered was an asymmetric method: its security has remained unbroken up until now. However, as we practically expected, a performance test of the RSA algorithm showed that this was not possible in our case. For the test, a very short stream of about 50 bytes was used. The decryption of these few bytes took almost a second. Within the real application, the amount of data sent is often more than one kilobyte. The time delay caused by this encryption and decryption would be unacceptable. Other asymmetric algorithms are as slow as the RSA algorithm, which also made them unsuitable for our client application on the PDA (Salomaa, Rozenberg & Brauer, 1996). An additional, very fast, symmetric encryption, which has not been patented, was implemented in order to ensure against a possible deactivation of the WPA encryption. There are many symmetric encryption algorithms. The One-Time Pad would have been best, but it is not practicable. DES and 3DES are old and also quite slow. IDEA, RC5, RC5a and RC6 are protected through patents. A5 is already broken. The result of this is that only RC4, Blowfish, Twofish and AES were suitable. Of course, there are some other (proprietary) algorithms; however, AES is a common and very fast algorithm.

This new standard algorithm has already been analyzed for security vulnerabilities, which practically ensures that there is no chance of cracking the algorithm within the next few years. Therefore, the AES algorithm was chosen. The implementation of the algorithm was not done within this project. A free implementation of the ACME Laboratories was chosen to encrypt the data. In this project, it is used with a 16 byte (128 bit) key which is calculated from a pass phrase which has to be specified in the property file. The used block size is 32 bytes, the maximum supported by this algorithm. This is acceptable because, although the calculation time needed for the encryption, decryption and the data lookup on the server is up to 5 seconds, the old list is left on display until the new one is loaded. It is possible that the customer might decide that no encryption on top of WPA encryption would be needed and change this in the property file to save time. It is also possible to change the whole algorithm, because this part was built to support other algorithms. The test with non-Standard English characters, such as the German "üöüß", was also interesting. These characters are encoded differently on the PDA. For example, on the PC the “ü” is represented by a byte with the value of -4. On the PDA, the value for the same character is -127. Other languages with additional character tables will have similar problems. To add support for these languages, an adjustment of the characters can be made after the transfer. Which changes have to be performed can again be specified in the property file which contains all the general settings. This makes it simple to adopt the system for many different character sets.

4 FRONTEND LESSONS LEARNED

At first the nurse logs on; the server then checks the logon information. The BSF communication framework provides the necessary security. After logon, the transport list is shown with the default filter. In order to provide the ability to change this, two option panels are available. Our main design criteria was to build the mobile interface as similar as possible to the existing non-mobile interface on the PC, which the nurses are used. We followed previous experiences on mobile interface design (Holzinger, Sammer & Hofmann-Wellenhof, 2006), (Holzinger, Searle & Nischelwitzer, 2007), (Nischelwitzer et al., 2007) and general experiences on usability engineering (Holzinger, 2005) in order to ensure an end-user centered user interface. On mobile interfaces, generally the keyboard size problem refers to the fact that using the on-screen keyboard significantly reduces the remaining screen size. The log on and the Options panel use Scalable Vector Graphics (SVG). The layout manager expects the whole screen to be available for the widgets. Constraints specify the place and the position as a percentage of the screen. This generally worked, due to the independence from the physical screen size. However, when the keyboard is selected, the program is notified to refresh its screen – to fit the
new, smaller screen; this new rendering causes each component to become smaller, which makes the options panel very difficult to read. After the removal of the keyboard, the PDA must again adjust the size of all components. This takes approximately 3 seconds, slowing down the overall application time considerably, which was perceived as extremely unpleasant by our test users. This problem can be solved by using another layout manager, however, the start up problem still has to be solved, where we followed a three step approach:

The first step is to review the needed libraries. Maybe not all classes of the library are needed. A new library should be built, which only contains the necessary classes. The reason for this is simple, to shrink the size of the libraries which have to be loaded, thereby decreasing load time. This is not very difficult, since the class files contain the information of their super class and interface, the dependency within a package can be discovered and the methods’ argument types can also be seen.

The second step is the loading process itself. This prototype loads all elements which are shown on the display at start-up. This means that all three panels, the logon, the transport list and the options, are all loaded before anything is painted to the screen. A better solution would be to load only the first panel, where the user enters their username and password. To do this, the user needs a few seconds to enter their account details: meanwhile the transport list can be loaded in the background. The benefit of this is simply that the user can start to work after a shorter loading time. From the user’s point view of while using the application, this does not make a big difference as the next panel is already loaded when it is needed. The same strategy can be used for the options panel. While the user looks at the screen and uses the functions of the transport list panel, the options panel can be loaded. The only disadvantage of this technique is from the programmer’s perspective: it is more work to load something in the background than it is to load everything at once. In this case, it is necessary to use separate threads to control which panels will be needed next and to check whether it is already loaded, or to load it before it is needed, in order to avoid delay. We had to ensure that the system waits until the required panel is loaded; in real-life this will happen very rarely.

The third step is to use separate frames for each panel. This helps to reduce the required memory, because after the user is logged in, the logon frame can simply be removed from memory.

The not default mistake means that it is default on the PDA that the close button in the title bar is used for the okay and cancel buttons. An X in the title bar means cancel and the symbol for okay is simply represented by the two characters ‘OK’. Because of this, the cancel button generally will not be used very often; and therefore could be removed to the title bar. Experienced users will be acquainted with this. Also, inexperienced users will not have many problems because they normally only need the OK button which is same as on the PC. When the cancel button is needed, a short glance over the rest of the screen will solve any difficulty that may arise.

The options scroll is really one of the bigger problems. This view shows the different options which the nurse can set up. As already described in the options paragraph above, every time the user activates a scrolling action the whole window must be redrawn. If there are only a few components on screen, then the PDA’s processor is sufficient to perform the necessary actions of the layout manager. Therefore, it is better to split the panel up into two windows, or panels, so that the scrollbar is no longer needed. A button is necessary to switch between the two panels. This makes sense because in this way it is possible to put the more important widgets on one panel and the remaining components on the other panel, which may not need loading as often. Another possibility is to reduce the number of possible options to a minimum by removing the less important ones and therefore the necessity for a second panel. Other known errors refer to usability difficulties rather than real problems, avoiding these makes it easier for the user to work with the program. A typical example includes the information displayed about the last update or a logout button.

The last update is necessary because if the nurse is somewhere in the hospital without a wireless connection to the server it is not possible to update the transport list. And because the nurse usually gets an update of the list every five minutes it is possible that they may think that the program does not work anymore because the list had not changed for more than ten minutes. In this case, the time of the last update indicates a connectivity problem, which requests that the nurse move towards an area with a wireless connection. This will mainly be an issue at the beginning, later support will be expanded to the entire hospital. The logout button is a time saving service for the nurses to enable them to share PDAs without the necessity of restarting the PDA application each time it changes hands.

When the nurse goes off duty they just log off and the next nurse can login immediately. Another thing which is not directly an error is the cancel button. Because this button has been moved to the title bar it is no longer a real widget which needs
space within a panel or window. So the okay button should be moved to the middle or to the right of the available space to fit the user’s expectations. Our last idea was to use different panels for the different screens and to perform any cancel actions using the “X” in the title bar. However, this was not possible. The reason for this is simple, that the default action on the PDA is to hide the window and to show the next one which is below the current window. On the PC it is possible to execute own code when the closing button is pressed. On the PDA this is not possible because no event is generated to perform an action which would prevent the window from being hidden. So some changes were necessary. The first was that every panel gets its own window, except the two option panels, which share one. The options window is the only panel which has a default cancel operation, performed by pressing the “X” on the title bar. This is possible because then the transport list window, which is the window below, is shown. The problem that switching the windows takes longer than switching the panels could also be eliminated by preloading the windows. If a window is needed, it is just shown on the screen because it is already in the memory, which prevents the shimmer. The other two windows no longer have the close button in the title bar to prevent user mistakes. Standard buttons are used to perform the needed actions. This means that the exit button, which was just removed, is again added to the logon panel. Also the logout button within the transport list panel can not be removed. In the transport list panel, one very important button is absent. There is no way to change the direction of the transportation. The patients have to be transported to the radiology and they must also be transported back to their stations. This means that two lists must be displayed. To save space on the screen the button is added instead of using one options button. The loss is that there is now no way to get directly to each of the options panels. To minimize the effects of this disadvantage the options dialog always shows the panel which was used the last time. A next/back button is used to switch between the two options panels. Other modifications which are needed are the logout and direction buttons and the change of the last update time string. The change of the two buttons has two reasons. The first is that the application should be consistent. In the options dialog, the position of the OK button to go back to the transport list is at the bottom left. The logout button is the same within the transport list window. The second reason is that the direction is like a headline. And so its logical position is at the top of the window. An extra window is used for the list, which is the main part of the application. In the top left of the window, there is a button to select the direction. The title of the button shows the currently selected direction. Possible values are “To Radiology” and “To Station”. By pressing it, the direction is changed. On the right of this button, there is a label which tells the user when the last update of the list was made. This is necessary because of the automatic update of the list. The time interval of the update is specified in a property file of the client. On starting, the list consists of the patients name, the time to transport, from where and if the patient can walk, which are displayed in two rows. After the selection of one of the items, the item pops out and shows additional information, such as the patient’s date of the birth. When one patient is transported, the nurse double-clicks on the patient in the list to mark this as transported. This is shown by the light gray background. After selecting another item the marked item pops in and the color of the font of the selected item turns dark gray. By the next update of the list, the update will have been completed on the server and the item is removed from the list. It is possible to declare a minimal time in the property file, during which an item is not updated to prevent mistakes, it is only necessary to make a second double-click on the marked item in order to change the item state back to normal.

Figure 1: Transport List: Before and after performing a transport.

When the button “Filter” in the transport list panel is pressed, the first options panel is shown. The nurse can specify different options for the filter; e.g. the transport type or the examination date. The transport list button behaves as a back button to the transport list panel; when pressed the expected panel is shown and the contained list is updated to fit the current filter settings. The “Next” button is used to get to the second options panel. The “X” is abort. The second panel is built as the first one. The whole available space is used for components which are used to specify different filter criteria. Examples for
filter options on this panel are the patient name or the handling of already completed transports. At the bottom of the panel there is again a button to return to the transport list and update it with the specified options to match the selected filters. The “Back” button switches to the first options panel.

5 CONCLUSIONS

It is not trivial to develop both a useful and usable mobile application. The general idea to use a J2SE Virtual Machine was abandoned, due to lack of compatibility. The server communication could not be done through the available methods. This made it necessary to build a relay to forward the inquiries to the server. That made it possible to convert the result to a string which is sent as a byte stream. Due to this sensible data, encryption was necessary. Primarily the WPA encryption is used, which is secure and is supported by the PDA’s hardware. Additionally, the data is encrypted by the application. After a comparison the AES implementation of the ACME Laboratories was deemed the optimal choice. Due to the fact that it is the new standard algorithm for symmetric data encryption, the algorithm can be graded as secure. The user interface of the PDA client also caused some troubles. It is not possible to build a slightly more complicated interface when using the default layout managers. It was necessary to program a custom interface to fit all our needs within an adequate time. Another problem was that the PDA had a different window handling. It is not possible to receive any window events within a Java application and that must also be considered. Another big problem on the PDA, which is not an issue on the PC, is the scrolling of complicated interfaces. The scrolling time is much too long because a simple scroll request required a few seconds of time to perform. Everything had to be optimized to save as much computing time as possible. However, some of the challenges of migrating desktop applications to mobile technology are the limited screen size and the originality of the technologies that need to work together correctly for a useful and usable solution. In conclusion, the PDA prototype can reduce the time needed for transports and the end-users were able to use the PDA similarly to their non-mobile interfaces.

REFERENCES

Brightsidefactory (2008), Bright Side Framework Overview. Online available: http://www.bs-

factory.org/components/remotingDoc/architecture.htm
l, last access: 2008-06-10
Freebeans (2008), Mysaifu JVM. Online available: http://www2s.biglobe.ne.jp/~dat/java/project/jvm/index_en.html, last access: 2008-06-10
1 INTRODUCTION

Many organizations use modern ERP systems for transaction processing and reporting. Information from applications within an ERP system is managed by a relational software database. Managers and executives need information for strategic and tactical decision that often requires the combination of data from ERP and non-ERP application sources. The ERP’s database complex schema is not adequate for data analysis and reporting. Business Intelligence technologies can provide the best solution for extracting, transforming and presenting the huge information from within the database.

The specific aspects may come from the way of implementing the BI project. In our case, the BI project was implemented by the same developer that implemented the ERP system, and it was strongly related to the ERP project implementation, which continued to be extended along with our BI project.

This paper describes a BI project experience developed in the context of implementing Oracle ERP solution into a big company activating in Romanian oil industry services, with a complex organizational structure, having over 15000 employees. We’ll analyze the difficulties and the problems met as the project was developed, and finally we’ll formulate some lessons learned of use for future implementations.

2 BI ENVIRONMENT: THE ERP SYSTEM AND THE BUSINESS REQUIREMENTS

In (Larissa T. Moss, 2001) we can find out that 60%-70% of business intelligence (BI) applications fail and these failures are not related only to the technology but to organizational, cultural, and infrastructure issues.

The ERP project apparently started on a solid ground, with Oracle Financial (FI) module implementation at the central unit. The ERP project then continued by implementing Oracle FI in all company branches, sectors and working points.

The implementation generated the need for processes reengineering and revealed important flaws of the company structure and management. In order to complete these tasks, many changes had to be made. These changes affected company structure, business information and material flows and the way they were registered by accountability, management...
and reporting hierarchy. It was also considered implementing the Logistics and Work over modules.

In this context, the BI project started when the financial module had already been implemented in the central quarters. In the beginning, the project seemed to be a classical one, and we followed the classical phases described in (Larissa T. Moss, Shaku Atre, 2003): Planning, Business analysis, Design, Construction and Deployment.

During project stages we took into account some specific in modeling BIS (Business Intelligence Systems), such as:

- BIS are oriented on business opportunities rather than transactional needs;
- BIS have to implement tactical and strategically decisions, not only departmental or operational decisions;
- BIS analysis is focused on business needs. This stage is the most important of the process;
- Development process is cyclical, focused on evaluation and improvement of successive versions, not only building and major delivering of a singular final version.

In order to respond to all business needs, the phases of the project were completed almost entirely twice: for the initial company structure and for the modified company structure. During this process, the BI team faced three important types of problems:

1. Technical Problems – the initially allocated resources were reduced, as the initial estimations of the data volume were overcome as soon as the BI module was implemented in the entire company. It was estimated that a virtual data warehouse would be sufficient, but the reporting and analysis needs were higher, so the performances of the initial solution had to be somehow improved.

2. Communication Problems – there was a lack of availability from the company personnel, as they were already overwhelmed by tasks related to the ERP implementation and by activities related to the consultancy companies’ cooperation.

3. Organization changes Problems – the BI project was from the very beginning under the sign of change. Beside the changes due to technical aspects, drastic structure changes in the company negatively influenced the relevance of the initial business analysis, and often made some of the developed reports irrelevant or inapplicable.

The work was performed with gaps, as there was always some top management decision to make that affected the BI project. The initial plan was revised over and over, so it somehow lost its power as management tool. The entire structure of incomes and outcomes categories and their corresponding accounts were changed in the middle of the project.

In the following sections we will focus on the technical and organizational structure aspects.

3 BIS’S DEVELOPMENT PHASES

The system described in this article includes the use of the following technologies: data warehouse realized with aggregate data and virtual model, OLAP technology, data mining facilities, analytic SQL functions. The interface is based on Oracle Business Intelligence Portal that integrates all these technologies and offers direct access to the system from anywhere, anytime.

First Phase

The cost factor was decisive when deciding which of two options to adopt:

- A traditional data warehouse, having aggregate data, involving an extended data volume, both elementary and aggregated.
- A virtual data warehouse, which worked directly with data from the ERP database, much easier to implement.

For this phase, we used Oracle Discoverer to build a virtual data warehouse over the database. Multi-dimension structures like dimensions and fact tables were transformed and mapped from relational sources. We started with low data volumes, generated by the use of the FI module at the central quarters. Oracle Discoverer Desktop was installed on the computers of the main decision makers that were going to use it. We developed reports for analyzing Cash Flows, Budgets, Costs, Expenses and Incomes only for the central head quarters. The number of rows involved in these reports was in average 100000/month, so we did not need a large amount of resources.

Second Phase

When the financial module was installed over the entire company, we faced the problem of having to go to each of the sectors to install the product and instruct the managers how to use it and some other IT persons how to ensure the maintenance and future development. In order to minimize the effort, we integrated Oracle Discoverer with Oracle Application Server Portal 10g Release 2, and the solution became available on-line, allowing everybody to access information according to their established privileges. But, this decision had a negative impact on performances: the working speed
decreased as the required memory space increased from 32 GB server shared as Discoverer needed, to 64 GB server shared as Oracle Portal needed. We didn’t manage to get a dedicated server for Oracle Application Server because of high hardware resource cost. In this phase, we had to develop reports for analyzing Cash Flows, Budgets, Costs, Expenses and Incomes for the central headquarters and ten sectors with different area of production. So, the number of rows involved in these reports was over 1, 5 million/month. Under the circumstances, with low allocated resources and a large amount of data per month we needed to find another solution to increase the speed and reduce the loading time for reports.

Figure 1: The system’s architecture in the final phase.

Third Phase
When the other ERP modules started to be implemented, we were facing serious problems. We decided to construct an enterprise data warehouse integrating Oracle Data Warehouse Builder and Oracle Data Mining into Oracle Portal. Unlike Oracle Discoverer, in Warehouse Builder data are physically stored into a special repository, build on the database. The role of Oracle Data Mining was to supply additional analysis power.

This solution was a success as data were aggregated in the central data warehouse, the ETL (Extract, Transform and Load) process was built off-line, not affecting the reporting analytical time. We built new types of reports related to production (costs and allocated resources), work over, purchasing orders, transactions per account, etc. We designed one central data warehouse with sub-data marts for financial, production, purchasing orders and order entry modules. The final solution is described Figure 1. The number of rows involved was now over 3 millions/month, and the performances were finally satisfying.

4 EVALUATION OF THE FINAL SOLUTION

4.1 Organization Changes Impact
Major business reorganization is one reason BI project typically get hosed (Mitchell A. Ocampo, 2007). In this project we faced an important business restructuring. More, the initial scope estimation proved to be inexact, and it had to be re-evaluated, so the project gained new dimensions.

Most of the analysis results had to be re-evaluated, and database design and repository design had to be revised and seriously extended, having an important impact over technical and human resource requirements. The BI team had to be enlarged and it was itself a problem, as there were no other BI specialists available, and other two new persons had to be employed and integrated.

Many of the reports developed in the first phase were inapplicable, generating important time loss and inefficiency; it often happened that after getting complete specifications, data gathering, and completing of the report design, during the validation meeting we were communicated the whole perspective was changed. This had impact not only directly upon efficiency, but our team members got to be confused or demoralized.

But, there was a bright side in all those changes. As the consultancy company progressed in elaborating the new business view, they inevitably brought out the reporting needs. Our team was invited for discussions, and we had the chance of proving the great advantages of BI over the existing reports. A demonstration based on the customized reports had a relevant impact.

There was also a change in the attitude of end-users, previously reluctant to changing their old tools (Microsoft Excel and other Office applications) and to learning how to use a new one. They were now eager and motivated as the old reports were no longer valid. And the truth is that they come to love it in very short time, as the new reports were very flexible and easy to use and they interfaced directly with the ERP application and other data sources.
The end-user involvement is a very important factor in developing functional BI solutions. As project requirements and scope are set by users, they know exactly the data they need and they will test and use the result of the project. There are voices claiming there was a shift of authority and ownership in BI projects, from IT to business side (Larissa T. Moss, Shaku Atre, 2003).

4.2 Criteria for Evaluating Business Intelligence systems

Deploying BIS involves many risks: system design, data quality, and technology obsolescence. Large budgets and strategic information are at stake in deploying BIS systems, a good reason to establish rigorous criteria for evaluating this type of system. These criteria are discussed below.

1. Decisions based on Business Process
BIS should not be viewed only as a data repository or a large set of data. Instead, system’s implementation should be concerned with conceptualizing new data models, processes, and indicators that form the content of the BIS.

2. Performance
It reports to the response time the system provides to its users, depending on the complexity of the database and the queries being requested. Most responses should range from few seconds to a maximum of 30 seconds for routine queries.

3. Flexibility and Scalability
Flexibility determines whether a BIS solution can adapt to changing business conditions after the system’s delivery. BIS should be able to accommodate changes in any type of business processes and functions, as well as in new laws and regulations, requiring the capture of different types of data. BIS should be expandable to accommodate data growth and changes in organizational structure, without a slowdown in performance.

4. Integration
Integration involves two types of issues: data integration and system integration. BIS will be particularly effective if it can overcome the challenge of information fragmentation, allowing managers to measure features of business processes that involve information from inside and outside of the organization. System integration involves two aspects: the ability to extent the BIS software with new capabilities and modules and the system’s ability to coexist with other enterprise solutions.

5. Friendly user Interface
BIS should be designed to provide managers a fast, easy, and understandable way to navigate into data and identify trends and patterns. BIS should allow the user interface to accommodate different degrees of technical knowledge.

We notice these criteria are related not only to technical aspects, but also to business aspects, as managers offer the most important feedback to the quality of business intelligence systems.

5 CONCLUSIONS

The most important conclusion of this real project experience is that, more than technical reasons, organizational and infrastructure dysfunction endanger the success of a project. When there are no supporting business decisions, the BI project is paralyzed. So, the main factor in conducting such a project is business rather than IT. Following this direction, after our project implementation was ready, the maintenance and further development of BI reports were entrusted to some end-users, specially trained (non IT).

Only combining BI technologies on the top of an Enterprise Resource Planning system will never provide a viable response to the changing management needs. A competitive advantage can be obtained only by adding business-oriented personalization so that the information provided can fulfill the particular needs of end-users and empower dynamic analysis and decisions.

REFERENCES

Larissa T. Moss, Organizational Barriers to Business Intelligence (Part 2), http://www.businessintelligence.com/ex/asp/code.135/pagenum.1/xe/article.htm
WEBSITE INTERACTIVITY
_e-Commerce Usability Perspectives in Indonesia_

Vincent Didiek Wiet Aryanto
_Graduate School of Business Management Soegijapranata Catholic University, Semarang, Indonesia_

vincent-aryanto@untka.ac.id

Keywords: Website, interactivity, usability.

Abstract: Usability is a key factor of website that is commercially successful. Interactivity has been linked to website usability. This research confirms that improving website interactivity promotes website usability. In addition, the extension of previous research such as usability sub construct such as learn ability, memorability, efficiency, user error and self-efficacy are validated to improve the body of knowledge particularly on e-commerce research.

1 INTRODUCTION

The use of e-commerce alleviates Indonesian textile industries nowadays that are beset by economic crisis. As a matter of fact, using e-commerce increases firm’s revenue up to twenty percent. Consequently, it increases firm’s efficiencies in production and marketing as well (Kompas, 2004). The impact of internet on marketing and/or advertising has been very important in current years, and it continues to proliferate (Ha, 2003; Macias 2003). The internet exposes to advertising, giving consumers the discretion to participate to particular messages within the medium (Klein, 2003). Marketing has changed dramatically due to the development of the internet (Strauss and Frost, 1999) and since its inception; online marketing including e-commerce has seen rapid change (Rowley, 2001). Growing online competition and maturation of internet technology have underpinned the expansion of website factors, beyond extensive product offerings, customer convenience, ease of navigation, and security, that affect online marketing success (Choate, 2000; Liu et al, 2000; Lohse et al; 2000).

Unfortunately, this is not true for most formats of internet advertising, advertising formats such as buttons, spams, banners, hypertexts, pop-ups; bars are not requested by the receiver of the messages. These advertisements are commonly incorporated in portals, search engines, or media websites, reach large audiences (Ha, 2003).

When the consumers is interested, he or she clicks on one of these advertisings and begin to control the communication process, hence, advertising exposure is able to active (Chaterjee et al, 2003). Websites, therefore, represent the most important form of interactive advertising. Websites are based on information and communication technologies enable easy and quick interaction between consumers and marketers (Coyle & Thorson, 2001). In a website, individuals can interact with the medium itself, which is called “machine interactivity”. The machine interactivity allows consumers to control what information will be presented, in what order, and for how long. (Ariely, 2000).

Website interactivity is a design factor that has been indirectly linked to website usability. Furthermore, interactivity is defined as mutual communication between at least two entities, with website interactivity generally involving communication between a user and the website (Liu & Shrum, 2002). Research on website interactivity is still in its preliminary stages, and there is a theoretical gap exists in describing and foreseeing the linkage between usability and interactivity. Seemingly, it may be intuitive that websites with a high extent of interactivity are more usable than less extent of interactivity, there is little theoretical and empirical proof to support this notion. So far, research indicates that interactivity can be useful, value-neutral, or detrimental to a website, depending on the situations. So, understanding how interactivity affects usability is essential to the
progress of research on usability in e-commerce websites. (Liu, 2003).

Usability literature provides guidelines for improving website usability; however, many of these guidelines are not supported by theoretical and empirical evidence causing them incorrect, and inapplicable in some contexts. Accordingly, many usability guidelines largely ignore interactivity aspects of user engagement. Thus, practitioner-based usability guidelines will not necessarily provide appropriate levels of website interactivity. (Chen & Yen, 2004).

In response to the increasing demand for e-commerce, it becomes more important to promote recent usability literature with research based on validated theory and empirical data. This research is based on the improvement of the previous research undertaken by Lowry et al; 2006. That suggested to take into account on usability sub construct such as learn ability, memorability, efficiency, error prevention and self efficacy as suggested by Liu and Shrum (2002), aside from satisfaction, this research. In response to the existing gaps in interactivity literature we set forth some research questions:

- Should perceived measures of two way communication, active control and synchronicity highly correlated with each other?
- Is there any positive correlation between website interactivity and website usability?
- Is there any positive correlation between website interactivity and website usability sub construct satisfaction?
- Is there any positive correlation between website interactivity and sub construct self efficacy of user?

2 LITERATURE REVIEW

2.1 Interactivity

Most of the definitions of interactivity particularly refer to two-way communication and interchange in which two or more parties send related message to each other (Liu & Shrum, 2007). Pavlik (1996) defines interactivity as a process of reciprocal influence (p.135), while Alba et al; (1997) proposed that two dimension of interactivity generally focuses on face-to-face interpersonal communication, however, Ha & James (1998) contended that the definition of interactivity based on interpersonal communication is too restrictive to fit technical communication. This assertion is particularly relevant when considering work that takes place over the WWW (World Wide Web) and related technologies. In the technology point of view, interactivity concept includes both human beings and computers as parties in an interaction.

2.2 Website Interactivity

Website interactivity is pondered to be a subset of interactivity that focuses on the interaction between websites and users. In this regard, website interactivity is broadly defined as any action a user or a website takes a time, while a user is predetermined end-user goal with the website, such as accessing information—whether users are seeking specific information or just browsing. Other goals may include business transactions or even entertainment (Stommer & Galley, 2000).

Interactivity of a website offers facilitated communications, customization of presented information, image manipulation, and entertainment for the customer. (Mathwick, 2002). The interactive nature of websites has been credited with increasing the desire to browse and purchase online (Li et al; 2001).

According to Liu and Shrum (2002) a two-way communication, active control and synchronicity mold the basis of a powerful, yet simple. Further development in two way communication refers to bi-directional flow of communication between communicators. (Liu, 2003). **Active control** is present when a user is given the ability to choose information and guide the interaction. **Synchronicity** indicates the timing of information exchange; a more concurrent or simultaneous exchange between two entities is a more synchronous interaction than an exchange with a longer time lag between responses. These three sub-constructs of interactivity form the basis for exploring the relationship between website interactivity and website usability.

The effects of fulfilled and unfulfilled expectations and desires of interactivity in websites are preliminary supported by website interactivity research. Wu (1999) found out a strong, positive correlation between positive perceptions toward a website and interactivity. Liu and Shrum (2002) provide a literature review and theoretical proposition to support the notion that two way communication, active control and synchronicity improve interactivity and communication with websites. Lowry et al; (2006) research’s to students taking Management Information System
coursework, supported Liu and Shrum’s theoretical proposition that two way communication, active control and synchronicity have strong positive correlation with website interactivity.

### 2.3 Website Usability

The more usable websites, the more positive attitude towards online marketing or e-commerce and increase repeated visit rates, while websites with low usability have the opposite effect. Nielsen (2003) defines usability in terms of learn ability, efficiency, memorability, user errors and satisfaction. Concerning website usability, Lowry et al; (2006) applied on his research, Nielsen’s theoretical framework, but only for the satisfaction variable, since the researcher believed that satisfaction triggered website usability and ultimately, satisfied users revisited designated websites. However, Lowry et al; (2006) suggested for future research to explore usability in terms of learn ability, efficiency, memorability, user errors. Lowry et al; (2006) used modified measure of satisfaction as developed by McKinney et al; (2002) and found out that there was strong positive correlation between website interactivity and satisfaction. The attribute of subjective satisfaction used by Lowry et al; (2006) refers to how pleasant a user finds it to use computer application or websites.

**Self Efficacy** influences people’s choice of activities, how much effort they will expend, and how long they will sustain effort in dealing with stressful situations. Self efficacy expectations toward the emerging traditional computer-based learning systems are likely to influence how participants use the systems (Bandura, 1996). Users with high self-efficacy tend to be more persistent in their learning and more confident in their ability to use the systems. Disorientation is one of the problems that novice explorers tend to have while navigating within a cyberspace (Dias et al; 1999).

Lowry et al; (2006) suggested for future empirical research the use of self-efficacy to measure website usability. Previous empirical research done by Liu and Shrum (2002) also suggested exploring users’ self-efficacy. We now operationalize hypotheses to test various elements of our theoretical extension of website interactivity to e-commerce website usability as follows:

**H1** The perceived measures of two way communication, active control and synchronicity should highly correlate with each other.

**H2** There is positive correlation between website interactivity and website usability.

**H3** There is positive correlation between website interactivity and website usability sub construct satisfaction.

**H4** There is positive correlation between website interactivity and sub construct self efficacy of user.

### 3 RESEARCH PURPOSE

This research entitled: Website interactivity: E-commerce Usability Perspectives aim at exploring the following:

- The perceived measures of two way communication, active control and synchronicity correlate with each other.
- The correlation between website interactivity and website usability.
- The correlation between website interactivity and website usability sub construct satisfaction.
- The correlation between website interactivity and sub construct self-efficacy of user.

### 4 RESEARCH SIGNIFICANCE

This study contributes to the theoretical understanding of how website interactivity relates e-commerce usability and its sub construct such as satisfaction and self-efficacy. Because user satisfaction, learn ability, memorability, efficiency, user error, and self-efficacy have been identified in defining website usability. Poor website usability may lead to the misgivings in e-commerce (Schneiderman, 2000). Rising interactivity shows potential method of promoting consumer trust in e-commerce websites via improved usability. A deeper understanding of the relationship between website interactivity and website usability provides several theoretical contributions by validating the theoretical models of Khalifa and Liu (2003), Liu and Shrum (2002), and extending research done by Lowry et al; (2006).

### 5 RESEARCH METHOD

We tested the website interactivity and website usability of website owned by Indonesian Association of Textile and Apparel Industries (A.P.I/Asosiasi Pertekstilan Indonesia) composed of twenty nine company members. The website can be found at http://www.textile.web.id, this association is an open membership association, willing to
bolster joint e-commerce on one website. So far, API website has been inviting new members to join. Consumers are able to surf various products offered by the association interactively. This website was chosen for three reasons (1) Textile and Apparel Industries is one of the backbone industries employing workers and generating fifth biggest income nationally (2) Website participants from both parties (consumers and producers) require interaction with the website to furnish the intended task (3) This website is unique since it is a joint-membership industries or industries association. The experimental design attributed website usability as dependent variable and website interactivity as the independent variable.

The participants were 31 students from a Management Information System class at Soegijapranata Catholic University in Semarang Indonesia. Participants were sophomore-level economics and business majors; 80% were female and 20% were male. All major protocols were followed e.g., informed consent. They were not asked to mention their name at the questionnaire in order to fulfill objectivity.

The participants were assigned to surf and explore the intended website at the campus computer laboratory for three consecutive days (4 hours times 3 days = 12 hours). They were assigned as “ghost shopper” or acting consumers that pretend to transact with the companies. Participants were asked to what they expected and decided from API or Indonesian Textile website. Each question was rated on 1 to 5 Likert-like scale.

6 RESEARCH FINDINGS

To test H1 - the perceived measures of two-way communication, active control and synchronicity should be highly correlated with each other-we used Pearson’s correlations. These correlations provided support for H1 that was positively correlated to each other. However, two-way communication was correlated with active control slightly (0.231), whereas two-way communication correlated with synchronicity (0.477) significant at the 0.01 level (2-tailed). Active control correlated positively with synchronicity 0.444 significant at the level 0.05 levels (2-tailed). The weak correlation between two-way communication and synchronicity can be described that users faced a major obstacle particularly the degree of synchronicity with respect to internet speed and the responsiveness of website officers/masters to deal with the e-commerce users.

To test H2 - there is positive correlation between website interactiveness and website usability. The result was positively correlated between interactiveness and usability (0.656). Correlation was significant at the 0.01 level (2-tailed) which fully supported H2.

To test H3 - there is positive correlation between website interactiveness and website usability sub construct satisfaction. The result was positively correlated (0.271) but in a less sturdy sense and less statistically significant.

To test H4 - there is positive correlation between website interactiveness and website usability sub construct self-efficacy. The result was positively correlated (0.466). Correlation was significant at the level 0.01 (2-tailed), which fully support H4. We performed Npar test, and the distribution was normal.

6.1 Discussion

This research extended e-commerce website interactivity theory by testing the relationships with website usability. We found that perceived interactivity sub constructs of two-way communication, active control and synchronicity were correlated with each other. This finding bolstered the previous research done by Khalifa and Liu (2003); Liu and Shrum (2002) and Lowry et al; (2006). We also provided extension of usability sub constructs in terms of learn ability, memorability, efficiency, user error and self-efficacy that was recommended for further study by previous researchers. We captured these ideas to carry out in our research and eventually, the result was website interactivity, significantly correlated with usability. However, contrary to the previous research done by Lowry et al; (2006), we found that website interactivity was not correlated with usability sub construct satisfactions.

These phenomena can be explained that technically speaking, network performance was lousy in the campus area and its surrounding; this caused slower internet speed especially during office hours. Users encountered with problems for wasting time. Dissatisfaction also occurred when the images appeared on textile website was not many and attractive. Apparently, the website designer purposely designed the minimum images and pictures to be posted in the website to enable faster download and upload time. Fiore and Jin (2003) found out that fewer images appeared on textile and apparel website e.g., 3 dimensions of virtual images, images of apparel examples dissatisfied the users.
Slow response from the webmaster or persons in charge to serve and entertain the users in fact, has dissatisfied them.

We came up in the study that website interactivity was strongly correlated with website usability sub construct self-efficacy. The interactivity of this textile website was related to the users confident to surf, obtain information and transact business on the website.

7 RECOMMENDATIONS & CONCLUSIONS

7.1 Limitations & Future Research

Limitation of this research is that although the increased interactivity benefits in our sample website, however, high interactivity is not necessarily always good. Sundar et al (2003) have suggested that there is point interactivity affects users negatively; this point may be based on the character of the individual. Future research shall explore the effects of high levels of interactivity on website users.

Another limitation is that our research only focused on textile website, with respondents who were college or university students. Future research should incorporate travel agents, flight agents, blogs and other types of websites etc and should also use research samples from populations other than university students.

Since we determined the designated purposely textile website for the respondents or participants in the interactivity research, for the future research we recommend to open the participants in opting their websites.

The shortcomings such as internet networks that decelerate the internet speed, lack of attractive images and pictures, unresponsive behavior of officer in charge should be taken into account; otherwise, user distrust will be prevalence to use e-commerce.

Limited number of participants or respondents was also the weakness of this research; therefore, future research should involve various internet users.

7.2 Conclusions

This research confirms that improving website interactivity promotes website usability. It extended e-commerce website interactivity theory by testing the relationships with website usability. We found that perceived interactivity sub constructs of two-way communication, active control and synchronicity were correlated with each other.

Furthermore, result of the research was website interactivity, significantly correlated with usability. However, contrary to the previous research done by Lowry et al. (2006), we found that website interactivity was not correlated with usability sub construct satisfactions.

REFERENCES


1 INTRODUCTION

Each terminal is connected to the central management and database server to transfer the in/out record and report any abnormal situation. For data flow, as those in/out records are confidential, a good protection to prevent from leaking information is necessary. Wired connection is the best way as the information can hardly be unauthorized accessed. The Hong Kong Immigration Department also has the authority to use dedicated line to build WAN connecting among different locations. That is, using CAT5 or optical fiber to connect from different passenger control points such as Lo Wu, Lok Ma Chau to the central database server, located in Wai Chai. Apart from that, there is a wireless network (WiFi) connecting to palms of inspectors who monitor the operation of e-channel. As the information transferred through the wireless network is not that confidential, it is justifiable to use it. In fact, WiFi has certain kinds of protections for data transmission.

In terms of the authentication, the most important part of the e-channel, a double identification is implemented, which contains verification of smart ID card and biometric verifying using fingerprint. A Public Key Infrastructure (PKI) is applied in the e-channel system. Both the private key and public key are saved in the smart ID card and the private key is encrypted by user’s fingerprint. When obtaining the fingerprint after the user inserting the smart ID card, data of the fingerprint will decode the private key and check whether the pair of private key and the public key is valid, so that it will be able to identify the authentication. The database of Hong Kong Immigration Department is one of the most confidential and important databases in the government. A large amount of in/out records has to be processed every day.

2 BENEFITS

First, the implementation of e-channel can save labor cost. The traditional counter needs one inspector for each counter to monitor people who go through the customs and check their identities. However, in order to help those people who have difficulty in passing through the e-channel, in practice, one inspector is needed to monitor 4 to 5 e-channels. Therefore, the number of inspectors to be employed can be reduced, which helps to save much labor cost.

The second benefit is that immigration inspection with e-channel is more reliable than using the traditional counter. In the traditional counter, the inspector will just check whether the ID card is valid or not, and then check whether the photo on the ID card is similar to the user’s face. They seldom check...
in detail whether the user is the owner of the ID card. The reason is that the validity of your ID card will be checked before you can enter the e-channel, and it will then check your fingerprint to ensure that you are the owner of the ID card. Besides, in the e-channel system, there is an infrared ray scan.

The third main benefit of implementing e-channel is that it can save the space and time for the customs service. Some space is needed for placing the computer, some is for the inspectors to sit and some is left for the corridor. However, the e-channel does not need any computer or inspectors sitting there, but only a corridor for people to go through.

Last but not least, the implementation of e-channel can improve the image of Hong Kong. Since not many countries have adopted similar systems, having the automatic e-channel implemented may give Hong Kong a superior representation as a more internationalized city, so that the image of Hong Kong can be improved.

3 PROBLEMS

The first problem is about the data security. In the traditional ID card, only basic information is provided. However, in order to match the new e-channel system and other new functions of Hong Kong smart ID card, extra information which is more important and confidential has been added. Thus, private information of citizens may leak out if the key is exposed.

The second problem is concerning about the fingerprint identification used in e-channel. Sometimes, a person may be blocked in the e-channel. The most possible reason is that some people may have problems in their fingerprints.

The final problem is the chance of chip damage. Theoretically, the card reader retrieves information by detecting the yellow chip in the ID card. If there are any damages in the chip, the card reader may not read the information properly.

To cope with this problem, Hong Kong government has already provided a cover for the card. Furthermore, a landing contact of the card reader instead of the friction contact is adopted. For card readers of friction contact, the contact part is fixed.

4 COMPARISON

US e-channel is designed for the application of the US e-passport, which was newly introduced. The United States requires that travelers entering the United States under the Visa Waiver Program must have an e-passport if their passports were issued on or after October 26, 2006. According to the relevant information provided on the website of United States Department of State, the US e-passport contains a contact-less integrated circuit, which is a 64 Kbit RFID chip.

Compared to the technology used in Hong Kong smart ID card and the chip reader in e-channel, the application of RFID technology in the US e-channel helps to speed up immigration inspections. However, there are several security concerns about the application of RFID in e-channel, such as skimming the data in e-passport, eavesdropping communications between the chip and reader, tracking user of the e-passport, and cloning the passport chip in order to facilitate identity theft crimes.

Basic Access Control (BAC) is implemented to minimize the risk of skimming and eavesdropping. A pair of secret cryptographic keys is stored in the chip embedded in the e-passport. When the reader attempts to read the information in the chip, it engages in a challenge-response protocol that proves knowledge of the pair of keys and derives a session key. Only if authentication is successful can the RFID reader access the data stored in the chip, so that Basic Access Control reduces the possibility of unauthorized access to the data.

However, the Unique Identifier (UID) can still be communicated with the reader in this process, which could theoretically allow the passport user to be tracked. The United States Department of State uses a Random Unique Identifier (RUID) to prevent the use of UID for tracking. Each time the chip is accessed, the e-passport presents a different UID which is not associated with the UID used in sessions that precede or follow the current session.

Cloning is another security concern that someone may copy the information in one chip and store it in another fake chip. The simplest way to mitigate this action is to verify that the data in the chip match the data presented in the e-passport, by checking the photos and biographical data, etc. Additionally, Public Key Infrastructure (PKI) has been introduced to automatically confirm that the identity of the
person presenting the e-passport matches the data stored in the chip and shown on the passport. Even though there have been many security protections in the application of the US e-passport and e-channel, the technology is still not mature enough. In August 2006, a security researcher Lukas Grunwald demonstrated the cloning of a European Union e-passport at the Black Hat and DEFCON security conferences in Las Vegas. (Martin, 2006) The EU e-passport uses similar RFID technology to the US e-passport. However, Randy Vanderhoof, executive director of the Smart Card Alliance, claimed that the data encoded in the chip is digitally signed and locked by the issuing nation, and could not be altered even if the chip was cloned. (O’Connor, 2006) Besides, these data are only basic information presented on the passport data page, and a digital photo. Even if the chip is cloned, there will not be serious problems except that the photo may be used for other purposes. (Reid, 2006) Nevertheless, people still worry about the security of e-passport and the RFID technology used in e-channel. As mentioned above, US e-channel adopts the RFID technology so that data in the chip can be accessed about 10 centimeters away from the RFID reader. Using Hong Kong e-channel, people need to insert smart ID card, and the reader will contact the chip to read data. This makes the immigration inspection in Hong Kong slower than that in US. However, since RFID technology has more security problems, protections for the US e-passport and e-channel are more than those in Hong Kong. In terms of the biometric identification, US e-channel uses face recognition while Hong Kong e-channel uses fingerprint. Comparatively, fingerprint is more accurate according to security experts (Kanellos, 2004). Last but not least, electronic visa may also be introduced for further convenience. When the visa is stored in the chip in electronic form, it will save time for the Customs officers to check. Costs especially labor cost can be saved. However, it seems not appropriate to introduce electronic visa until security protection technologies become mature enough.

5 CONCLUSIONS

From the above, it is clear that various IT applications in Hong Kong e-channel have brought lots of benefits to citizens and the Hong Kong government. However, problems and concerns still exist. Therefore, Hong Kong government may constantly revise the system and learn from other countries for further improvement. E-channel is only one of the IT applications of Hong Kong smart ID card. More functions, such as public library service, can be performed with the smart ID card. The development of e-government has offered citizens and businesses quicker and more convenient access to government information and public services. There is a future trend to achieve a more accessible, accountable and efficient government for Hong Kong as a leading digital city.

REFERENCES


THE PRIVATE AND PUBLIC PARTNERSHIP STRATEGY IN E-GOVERNMENT ESDLIFE IN HONG KONG

Tak Ming Lam
Hong Kong Polytechnic University, Kowloon, Hong Kong
mslam@polyu.edu.hk

Keywords: Partnership strategy, e-government, ESDlife.

Abstract: Increasingly e-governments are turning to public and private partnerships as an alternative option to fund IT projects. Also The Hong Kong Government developed ESDlife.com as a one-stop portal to deliver public and private information and services. The portal’s brand name “ESDLife” best demonstrates the vision to make “life” simpler and easier for the Hong Kong community. This paper is to discuss whether this business model would deliver a competitive advantage to customers using the case of “ESDLife. E-government, like e-commerce, will not be automatically successful by investing resources without clear strategic understanding of the critical success factors.

1 INTRODUCTION

The ESD Scheme was launched in December 2000 as a flagship e-government project and its portal is one of the world’s first bilingual (English and Chinese) one-stop portals providing integrated electronic public and commercial services. The main aims of ESD are: to deliver high quality public services to the community in an innovative manner; to improve the efficiency and reduce the cost of delivery of public services; and to foster the development of electronic commerce in Hong Kong. Providing an open and common information infrastructure for the delivery of public and commercial services, the Scheme drives the development and adoption of e-commerce by the business sector and the general community.

Users can search for services through three mega-channels – “People”, “Business” and “City”; through nine service types “Transport”, “Citizenship”, “Education”, “Employment”, “Finance”, “Household”, “Leisure”, “Business” and “Tourist”; through the list of departments and agencies providing the services or through interactive search. Services are available 24 hours a day and seven days a week. Services can be accessed via personal computers with Internet connection; public computer facilities installed in district offices, community halls, post offices and public libraries; or smartly designed public kiosks installed throughout the city in train and subway stations, shopping centers, supermarkets, cultural and exhibition centers and Government offices.

2 ESDLIFE BUSINESS MODEL

The private sector operator is responsible for the development, operation and management of the ESD information infrastructure. The Government has not paid for the capital investment nor the recurrent operational and maintenance cost. It will only start to pay for each chargeable transaction made through the system after the number of cumulative chargeable transactions conducted has reached an agreed volume. The operator is allowed to make use of the system to provide advertisements and revenue-generating private sector e-commerce services via the same portal web site. This provides an additional income stream to the operator.

To the public and government, it can embark on such partnerships for the greater good, especially if the agency has the mandate to promote ICT adoption to domain industry partners, through leadership in ICT exploitation, to raise the overall competitiveness of industry or community specific services. Also the public agencies are no longer required to front substantial capital development and yet co-share in
e-services revenue collected. In addition if the public agency is not a dominant player to other parties in the value chain, then the agency could collaborate with other agencies so that the government as a whole becomes a dominant player.

From the private service companies views, we can see that the motivation lies on the potential additional profit stream and an opportunity to gain market share and domain knowledge of the public agency’s user base. The private sector is able to get the incentives to create more innovative and value adding services for the end users. And it has greater chance to build up goodwill in front of the citizens and the government so as to keep a long-term business relationship with the government.

Probably the most important success factor in implementing e-government in Hong Kong is the vision of the HKSAR Government. The e-Government Coordination Office was established to plan and oversee the e-government policies and initiatives, including ESDLife project.

3 ADVANTAGE OF THIS MODEL

3.1 Reduce Risk of Government

Under this partnership approach, the business risk to Government will be kept to the minimum. The financial charging model also provides an incentive for the operator to drive the public in accessing e-government services, continuously to upgrade the system with the most advanced technologies and to deliver private sector electronic services.

3.2 New Platform for Commercial Opportunity

More electronic business opportunities are provided by the government, and it could then enhance the e-business environment in Hong Kong. The scope of ESDLife requires the exploitation of new technologies, and it would then strength the community for digital exploitation.

3.3 Integration of Public and Commercial Services in the Same Portal

This enables the public to enjoy electronic services in a seamless and convenient manner. By allowing the ESD information infrastructure to carry commercial services, the project also promotes the development of e-commerce in Hong Kong. The e-government services and the commercial services can drive the overall adoption of e-commerce in the business sector and the wider community.

3.4 Quick Receive and Response to Customer Feedback

Apart from that, the ESD operator has carried out customer surveys to obtain customer feedback on the usage of ESD services. Such feedback is also important to improving the services under the Scheme.

4 DISADVANTAGE OF THIS MODEL

4.1 Registration Problem

Although the e-government initiative offered convenience to the public, there remained some inconvenience, in particular in relation to those electronic services that required an electronic certificate to allow the user to register on-line.

4.2 Third Party Software Needed

In addition to the electronic certificate, users needed to download and install special software in their computer. Two programs were required, one for security issues and another for the use of Chinese characters.

4.3 Connection Problem

Although ESD was made available in a variety of ways – in kiosks at around underground train stations and other public places – the system only efficiently supported users with broadband Internet hook-up, which meant that users who were connected to the Internet through telephone lines faced long delays in transmission.

4.4 Complexity of Process

For example, renewing a driver’s license through the Internet required scanning and transmitting an Identity Card and other documents, a process that took several minutes and equipments to complete.
This process was not considered simple, particularly for non-technical people.

4.5 Privacy and Security Issues

From the citizens’ point of view, the major concern for them towards the involvement of the commercial sector in providing e-government services is infringement of the personal data. It is true that more use of the personal information means more privacy concerns. It is possible that in collecting, transferring and accessing citizens’ data through the ESDLife systems for registration or transaction, the private sector may be inappropriate to use and even disseminate them without citizens’ previous consents. And lack of adequate control mechanism in reviewing private institutes would even bring this problem to a large extent. On the other hand, privacy will also surface as it is possible for the government to track online activities like frequently accessed products, the length of time spent on each page, and the length of time spent searching. While this tracking information can be used toward improving the website and its offerings, the temptation to sell this information to external parties may also exist since a close relationship with each other.

5 FURTHER IMPROVEMENT SUGGESTIONS

To develop a consolidated mix government portal, and to keep improving and well prepared for the challenge in future, we proposed the following suggestions:

Smooth Transformation from Traditional Processes to Modern Technology. Despite that the Scheme has offered a substantially more efficient and effective alternative means of service delivery, it is not a easy task to use the Scheme to replace the existing forms of providing services to the public over the counter or over the phone.

Privacy Enhancement. To alleviate the concern for the customers over the public and private sector collaboration in delivering government services, a clear and transparent guideline should be drawn to illustrate the liabilities and obligations between both parties.

Digital Signature Fee. Since digital signature is one of the methods used to provide data integrity, most of transactions occurring in the ESDLife need to adopt the digital signature technology to verify the users. However, in most of the situations, citizens are needed to charge $50 per year for the use of the digital signature.

User Interface. Firstly, ESDLife system should always inform the users what is going on. It is very important in payment status. Secondly, It is important to lead users to have more control and freedom in using the system.

REFERENCES

ESD Services limited www.esdlife.com
Information Technology and Broadcasting Bureau, Government of the Hong Kong Special Administrative Region, August 2001. “Hong Kong Report on Portals”
Jessie Ting, Deputy Secretary of Information Technology and Broadcasting, Government of the Hong Kong Special Administrative Region, December 10, 1999. “Impact of ESD on e-Commerce Development in Hong Kong”
A WEB-BASED ARCHITECTURE FOR E-GOV APPLICATION DEVELOPMENT

Marcelo Tilli, André M. Panhan, Osman Lima and Leonardo S. Mendes
Communication Networks Laboratory, Communications Department, Eletrical and Computer Engineering Faculty
University of Campinas, UNICAMP- Cidade Universitária “Zeferino Vaz”- Barão Geraldot, Campinas, SP, Brazil
marceltill@gmail.com, apanhan@gmail.com, osman_lima@yahoo.com.br, lmendes@decom.fee.unicamp.br

Keywords: Electronic Goverment, e-gov, web-based system, MVC, Hibernate, Struts, Service-oriented.

Abstract: In this paper we present a web-based architecture for e-gov application development. We propose to study and characterize a software development environment of “electronic governance in cities”, with emphasis on: legacy system integration, heterogeneous database integration and development of an integrated web-based environment for the provision of e-gov services. This architecture is based on a 4-layer MVC model, the typical MVC (Model, View, Controller) plus a data layer, this architecture imposes few requirements to its clients, mainly a HTML web-browser, which allows this architecture to be platform-independent.

1 INTRODUCTION

A great deal of the development effort of software systems is centered on modeling, characterization and development of business models for the solution of specific problems. However, when you have to deal with the development of e-gov systems, particularly when the systems are built in an integrated network environment, there is a great dependency among all the systems that compose the general solution of e-gov applications.

Systems of people registration, services, taxes, protocol, finance, are common on a government environment. The problem is that usually, several different vendors, which use different platforms and different development tools, often build these applications. Then we end up with several heterogeneous and independent data bases, which are often incompatible. This scenario is very common on e-gov systems around the world.

On the last ten years, several countries have sponsored e-gov systems and projects of system interoperability. Initially, these architectures were created to access specific problems and not with the objective of being a generic model that could be universally accepted.

There are many research areas on the topic of electronic governance systems (Peristeras, Tarabanis, 2004), (Narasimha, Kumar, 2003), (Murthy, Kumar, 2003), (Song Gang, 2005), (Lenk, Traummüller, 2001). The study of semantic information processing platforms, the development of interoperability platforms and legacy system convergence, are some of the many research areas that the study of advanced techniques of software engineering, computation engineering and communication engineering can contribute to solve several difficulties that faces e-gov system’s software developers.

In this paper we present a web-based architecture for e-gov application development. We propose to study and characterize a software development environment of “electronic governance in cities”, with emphasis on: legacy system integration, heterogeneous database integration and development of an integrated web-based environment for the provision of e-gov services.

2 PREVIOUS WORKS

The work of (Peristeras, Tarabanis, 2004). This paper summarizes a broad research modeling effort, which aims at developing a domain description for the overall governance system. The authors propose the Governance Enterprise Architecture (GEA) as a set of domain models that serve as a top-level enterprise architecture. Namely, they present the mega-process model, the interaction model, the public policy formulation object model, the service provision object model and the latest development of the object model for the overall governance system.
The work of (Narasimha, Kumar, 2003) propose an architecture for model development, for the planning, design and implementation of e-gov systems. The proposed model has four conceptual layers: Business Process Architecture; Data Architecture; Application Architecture; Technology Infrastructure.

3 SYSTEM ARCHITECTURE

We propose an architecture for e-gov systems for cities, that can access the general conditions of such systems, such as: the existence of legacy software, redundant, heterogeneous data bases and the paradigm shift from client-server platforms to web-based, integrated system for e-Gov services.

3.1 Architecture

For the architecture development, we used a modified MVC model (Model-View-Controller). In pursuit of greater security and flexibility, we separated the client’s communication layer from the business logic; hence we created an application layer responsible for dispatching requests and controlling its flows. The 4-layer architecture (Model-View-Controller-Data) is show on Figure 1 (Fowler, 2003).

This 4-layer model allowed us to reach some objectives: Facilitates the development of a distributed system, with security and with high availability; Enhances the automatization capacity of the city’s public administration sectors, offering online services to the citizens, expediting its execution and accomplishing greater overall quality.

This architecture imposes few requirements to its clients, mainly a HTML web-browser, which allows this architecture to be platform-independent. With this architecture, there’s no need to install/update the application software on the user’s computer. Modifications of the system’s software are made directly on the main server, which is transparent for the user. This is an improvement over the 2-layer model, where all update would have to be physically installed in the user’s computer (Fowler, 2003).

The 4-layer model is as follows:

3.1.1 Model (Business)

This layer contains all the business logic and manages this logic through the use of EJB components. Therefore, this layer provides a structure for the implementation of distributed applications, aiming for the separation of concepts and to improve on quality factors, such as modularity, extensibility and reusability (Alur, Crupi, Malks, 2001).

3.1.2 View (Presentation)

This layer renders the model into a user interface element. It access enterprise data through the Model and specifies how that data should be presented. We are developing this layer with JSP (Java Server Pages) and JSF (Java Server Faces)

3.1.3 Controller (Application)

This layer processes and responds to events, typically user actions, and may invoke changes on the model. Based on the user’s interactions and the outcome of the model actions, the controller responds by selecting an appropriate view. To facilitate the model’s utilization, this layer should be able to interact with default Data-Base access technologies, such as JDBC and EJB and also be able to interact with third-parties technologies such as Hibernate, iBATIS, or Object Relational Bridge.

3.1.4 Data

This layer is the domain-specific representation of the information on which the application operates. It
is responsible for the DAO (Data Access Objects), or the direct Data-Base access. This model allows for a transparent persistence of the objects models.

The data persistence allows us to aggregate several features to the proposed architecture, like the following: Transitivity persistence; transparent retrieval of data-objects (on a given object graph);

The purpose of the data-layer is to allow for an intelligent way to store object-oriented data, allowing for the convergence of heterogeneous data-bases.

4 AN E-GOV SYSTEM FOR CITIES

On this section, we will present a use case for a complete system for cities management, which can manage all services, citizen records, processes management and relevant data for a city’s administration. The main purpose of this system is to provide consistent information to the manager to make the right decisions. From this prerogative, emerged the SIGM (Integrated System for Municipal e-Gov).

The application integration under a same domain was possible with the development of a single data base, covering, for example, citizens, business, social data, departments, processes, internal services, online services, systems users, addresses and others. Therefore, all data are concentrated and correlated under the same concept “Citizen Single Data Base” (BDUC).

As this technology works with data object’s concept, the conventional relational data base BDUC was created to facilitate the implementation and take advantage of all of the Hibernate resources. Therefore, we created a data base under two perspectives: object-oriented and relational, thus, it is possible to represent any data evolved with e-gov and its legacy systems.

This modeling, as in object oriented system, allows us to reuse data for creation of new tables, without affecting the system operation.

In particular for this use case, we used the Microsoft Windows 2003 Server operational system and the Oracle DBSM, but we can also apply the same use case to others platforms.

The SIGM can be deployed in any city, even those that have legacy application in use; since this kind of integration was planned by the adopted architecture and can be done in several ways:
- SIGM control the legacy applications through its functionalities and the BDUC data base.
- The legacy applications are incorporated by the SIGM.
- The legacy applications can use the business rules already implemented in the SIGM’s EJB layer.

After deployment, the SIGM can be accessed from any computer in any city location, by any person (citizen or government employee) registered in the system. These users can request certain services through the Internet, without the need to move to the city hall or any other service point.

This architecture also offers the possibility for application distribution on multiple servers, not only in the system execution, but also on its development. By distributing the development of a system of this magnitude, we obtain independence between the developer’s teams and we can control more effectively each module implementation. By separating the business layer from the web layer, it is possible that others systems can use its functionalities without affecting the SIGM’s operation. This can be done by direct access of the EJB containers which are on the business layer.

To keep in order with the proposed architecture, on the web layer we adopted the Struts framework, which is an open source framework that works on MVC architecture and provides several control components to create dynamic Java web applications (Husted, Dumoulin, Franciscus, Winterfeldt, 2003).

This architecture also provided the development of an authentication system, which allows the SIGM to manage the permissions and profile control of the SIGM’s users and users of any other legacy
application already in operation, as shown in Figure 3. In these systems, all users and applications are registered through SIGM, and each one has its profile registered in different tables, so we can create different permissions profiles to different users for different applications. To control a legacy application, we just need to register the profile of this application and their permissions.

Figure 3: Legacy System Integration.

This use case was useful to demonstrate the efficiency of the proposed architecture and the technologies chosen for its implementation. In regard to this, we can consider that the use case was successful in its implementation.

The first version of the system was implemented and it is currently in use in the Brazilian cities of São José do Rio Preto and Campinas. We continue the system development, with new functionalities and features being added frequently, without hindering the system’s operation.

5 CONCLUSIONS

This paper presented an architecture (SIGM) for the development of e-gov systems, taking in consideration government environment such as: the existence of legacy systems, redundant and heterogeneous data bases, the paradigm shift toward web-based platforms, the necessity of constant services development by the government, etc. The architecture is based on the MVC model with the addition of a data layer.

The SIGM purpose is to integrate heterogeneous applications under a distinct domain, to develop a single data base, that encompass data from all the citizens, enterprises, social, government departments, government procedures, internal services, online services, system’s users, address, etc.

By using this single data base, all the data are converged and correlated under the concept of the BDUC, which data access is made through Hibernate, which allows the independence from a DBSM. SIGM can also control legacy applications through its functionalities and the BDUC data base.

REFERENCES


PROSPECTS OF GRID IN THE CURRENT VOTER REGISTRATION SCENARIO OF BANGLADESH

Sazia Mahfuz
Department of CSE, East West University, 43 Mohakhali C/A, Dhaka-1212, Bangladesh
saziam@gmail.com

Keywords: Grid approach to voter registration, Distributed election process.

Abstract: Bangladesh is a developing country which has gained independence in 1971. In the past, voter registration was done by going from house to house and then filling up the required forms and counting them in. So any type of computerized information system was not used. Currently, the government has taken initiatives to collect voter information with photograph to effectively execute voter registration. This is definitely a positive initiative from the part of the government. In this paper, to effectively limit the time duration and complexity of the ongoing process, grid computing infrastructure implementation is proposed. Grid infrastructure will not only accelerate the voter registration process, but also speed up the election process on the whole.

1 INTRODUCTION

The bangladesh election commission (BEC) is a constitutional body responsible primarily for the task of conducting elections for the parliament and the local bodies (http://www.ecs.gov.bd). A significant part of the task is the preparation, maintenance and periodic updating of an electoral roll conforming to the norms of the structure.

Recently the BEC faced significant difficulties in discharging its responsibilities due largely to the process adopted in preparing the electoral roll. The 9th Parliamentary Election was supposed to be held on 22 January 2007 following the dissolution of Parliament in October 2006. But, in the days leading to the elections, violence became widespread with deterioration in overall law and order situation of the country. Eventually, the elections could not be held. A new caretaker government assumed charge and vowed to the nation its intention of holding a free, fair and credible election.

One of the main reasons for the deferment of the January 2007 election was a faulty and grossly inaccurate electoral roll that was not acceptable to the opposition parties. To improve the situation, the newly reformed BEC took steps for preparing a credible electoral roll with photograph. The attempt has been appreciated to all parts of the people of the country. In this paper, an approach to speed up the current process further has been proposed.

2 VOTER REGISTRATION IN THE PAST

After the independence, Parliamentary elections were held for the first time under the new constitution on 7 March 1973. Later in the year, elections to local bodies were also held (“Banglapedia”,n.d.). According to the article no. 118 of the constituency, there shall be an Election Commission for Bangladesh supervising the whole process of elections. Up until now, eight parliamentary elections were held. In all of the election processes, the voter registration process was somewhat manual in the sense that the listing was done by going from door to door and keeping track of the voters in a written list. There the enumerators designated by the Election Commission take account of all the voters under his area. He hands down the list to his supervisor, who after further scrutiny submits it to the Assistant Registration Officer (ARO). Then the list is counted and the proper list is maintained in every polling centre.

This process had several drawbacks:
• A voter can cast multiple votes in different polling centres under different names as there weren’t any biometric identification.
• Fraudulent voter entry can be created due to the lack of proper identification.
As a result, outcome of the election process had become unreliable and fraudulent. General people had lost trust in the election process due to the inefficient techniques adopted for the voter registration process. This led to the unrest in January, 2007.

3 VOTER REGISTRATION IN THE CURRENT SCENARIO

An improved system has been achieved by developing a system whereby the computerized database is generated and is proposed to be maintained through annual update using a process of addition, deletion and transfer. This exercise involves the capture of photographs and the integration of these photographs with data in the computers. The whole process can be summarized through the following phases:

3.1 “First Phase: Data Collection”

Enumerators are provided with the appropriate forms and required items using which they collect data from the prospective voters. They have to keep records of the mentally or physically challenged people as well. The information is then handed over to the supervisors and after scrutinizing the data, it is handed over to the Assistant Registration Officer (ARO). The data is compiled according to the areas after scrutiny by the ARO.

The voter registration centres and area wise server stations are then set up with assistance from army. The registration forms are provided to the team leaders by the AROs. The team leaders then inform AROs of the data, time and numbers for photograph taking. The information is then again passed on to the enumerators from the AROs through the supervisors. The enumerators inform the prospective voters about the time and date on which they have to be available in the voter registration centre for photograph taking. The necessary preparation is taken in the voter registration centre, i.e. the laptops, webcams, fingerprint scanners, photocopy machines and generators are set up for the operations to be performed for data collection.

3.2 “Second Phase: Re-scrutinizing the Forms”

The forms maintained at the registration centres are scrutinized again and then entries are created in logbooks after serializing the forms in laptops.

3.3 “Third Phase: Data Collection and Identity Scrutiny at Registration Centres”

Voters provide their information to the data entry operators after scrutinizers have verified them. The entries are created if required and completed after taking the fingerprints, photos and signatures from the voters. Then the voters are given receipt.

The information thus procured is again handed over to the team leaders using laptops and they are transferred to the area wise server stations.

3.4 “Fourth Phase: Registration of Physically Challenged, Jail Inmates, Missed-out Voters and Ailing People”

After collecting data from the regular voters, all the other prospective voters who have missed out the registration have to be taken into account. They have to be counted for following the steps described in the earlier phases. The registration for the physically ill people and the jail inmates have to be completed as well, under the required arrangements.

3.5 “Fifth Phase: Data Processing at Area Servers”

As all data have been collected for the voters, the information has to be processed in the respective area servers. The duplicate voters have to be identified and the team leaders are informed. Then the errors are corrected as required. The national ID cards are processed for distribution after identifying and correcting the errors in the exhibited draft national ID cards.

3.6 “Sixth Phase: Draft Voters List Printing and Finalizing”

Finally, the draft voters’ list is prepared and exhibited. After hearing the objections on the list, corrections are made by revising the proper authority. Then the final voters’ list is prepared and authenticated by the registration officer.
The electoral roll with photograph thus prepared is proposed to be merged into a central database by incorporating the fields necessary for voter registration. This initiative has been laudable compared to the earlier system that had always been used for voter registration. But will it be successful or not, that is yet to be seen. A better approach to implementing this scenario will be grid computing infrastructure.

**Figure 1:** Physical movement of information in the fifth phase of the existing scenario.

### 4 VOTER REGISTRATION IN THE PROPOSED SCENARIO

The ongoing voter registration can be further improved with respect to the required time. The proposed plan is to implement grid infrastructure for the voter registration process. Currently, the voters’ information has to be transferred to the area servers from the voter registration centres for data processing as described in the fifth phase. Our proposed plan suggests a change of the architecture used in this phase.

The voter registration centres already have laptops, fingerprint scanners, webcams for taking information from the voters. Every voter registration center is also connected to the existing public switched telephone network (PSTN) provided by BTTB (Bangladesh Telegraph and Telephone Board) (http://www.bttb.net.bd/). So if there is an online interface for connection to the grid computing system (Foster, 2006), then the need for the transfer of information from the registration centres to the area wise servers will be eliminated. All the voter information taken in all the voter registration centres can be directly fed into the proposed grid infrastructure.

This approach will obviously limit the time duration required for the data transfer into the area wise server and data processing required in those server centres.

Moreover, it will pave ways for many more opportunities for e-government related services.

**Figure 2:** Proposed scenario for Voter Registration in grid implemented approach.

### 5 ANALYSIS

According to the roadmap (http://www.ecs.gov.bd) provided by the Bangladesh Election commission, the process of enumeration, data entry, printing, posting and correction of the draft list will start from 7th August 2007 and finish on 16th August 2008. So this phase of the process will have an approximate time span of one year. This extended time duration is due to the lengthy process of data collection and then transferring the information to the area wise server for re-evaluation.

This time span will be substantially reduced if grid infrastructure is implemented. A simple online interface can be developed to use the distributed access points. The voter registration centres distributed all over the country can be used as the access points for the grid. The components required to maintain the grid can be distributed in any number of servers located at the different government bodies secured from outside intervention. The need for manpower in local area wise server maintenance can be removed through this approach.

### 6 RISK ASSESSMENT

“A gap exists for all e-government projects between the design assumptions/requirements and the reality of the client public agency. The larger this gap between design and reality, the greater the risk that the project will fail (“DRG”, n.d.).” The technique
followed in design-reality gaps model ("DRG", n.d.) presents a set of questions which will attempt to rate the gaps.

The DRG model involves questions relating to a series of seven ITPOSMO dimensions - information, technology, processes, objectives & values, staffing & skills, management systems and structures, and other resources - with attached rating numbers ("DRG", n.d.). Using each of the seven dimensions, two things have to be observed. One is the organizational reality that exists at the time of analysis and the other one is the requirements within the design of the proposed e-government project. For each one of the dimensions, a numerical rating has to be allocated to indicate the size of the design-reality gap on that dimension. The rating for each dimension’s gap can be anywhere on a scale from zero to ten.

After the rating, the rating numbers for all seven ITPOSMO dimensions are added up and can be interpreted according to the table 1 ("DRG", n.d.).

The seven dimensions are applied to the proposed approach and the corresponding rating is discussed:

- Information: The information currently being used for voter registration is the same as required by the approach. So no change indicates 0 rating.
- Technology: The required software for the development of the grid computing environment would be slightly different than the setup currently being used. So a rating of 4 can be allocated to this dimension.
- Processes: The work processes undertaken in the current scenario need not be increased for the requirement of the proposed scenario. So a rating of 0 can be assigned to this dimension.
- Objectives & values: The objectives and values required for the key stakeholders of the proposed system have to be exactly the same as the objectives and values of the current system. So a rating of 0 should be assigned corresponding to this dimension.
- Staffing & skills: The staffing members and the skill level or type will slightly differ in maintaining the control of the grid architecture. So a rating of 3 has to be assigned for these criteria.
- Management systems & structures: Management systems and structures required for the implementation of the proposed approach will be slightly different than the exiting system. But the overall management will be the same. So a rating of 4 can be assigned for this dimension.
- Other resources: The time and money required to implement the proposed system will be higher for the initial setup. But it will definitely be lower for maintenance compared to the existing system. So a rating of 2 can be assigned.

The total rating for the ITPOSMO dimensions is added up to 13. According to the table 1, there is a high possibility of success in the proposed approach.

<table>
<thead>
<tr>
<th>Overall Rating</th>
<th>Likely Outcome</th>
</tr>
</thead>
<tbody>
<tr>
<td>57 – 70</td>
<td>The e-government project will almost certainly fail unless action is taken to close design-reality gaps.</td>
</tr>
<tr>
<td>43 – 56</td>
<td>The e-government project may well fail unless action is taken to close design-reality gaps.</td>
</tr>
<tr>
<td>29 – 42</td>
<td>The e-government might fail totally, or might well be a partial failure unless action is taken to close design-reality gaps.</td>
</tr>
<tr>
<td>15 – 28</td>
<td>The e-government project might be a partial failure unless action is taken to close design-reality gaps.</td>
</tr>
<tr>
<td>0 – 14</td>
<td>The e-government project may well succeed.</td>
</tr>
</tbody>
</table>

7 CONCLUSIONS

Grid computing has become the popular solution to distributed problems. Election process is a very critical distributed problem faced in any government. In this paper, a better approach is proposed to maintain the database of the voter registration, so that the distributed election process can effectively be handled with the least possible use of centralization.

REFERENCES

WEBSITE CREDIBILITY
A Proposal on an Evaluation Method for e-Commerce

Katsuya Watanabe, Masaya Ando and Noboru Sonehara
National Institute of Informatics, 2-1-2 Hitotsubashi, Chiyoda-ku, Tokyo, Japan
charley@nii.ac.jp, ando@nii.ac.jp, sonehara@nii.ac.jp

Keywords: Information credibility, website credibility, web trust.

Abstract: This paper introduces a research for clarifying the structure of the website credibility. The users with low information literacy cannot have difficulty making use of e-Commerce services because they cannot judge the credibility of websites appropriately. Conventional approaches to evaluate the credibility of a website have been based on aspects like the design of the website or the usage of information security technologies. However, it is not sufficient for users with low information literacy to distinguish ill-intentioned sites based on the site design and security technologies alone. In this paper we examine a more comprehensive analysis and evaluation method which is based not only on evidence internal to the site, but also on third-party information about the site provider.

1 INTRODUCTION

Access to and use of the Internet has spread widely and swiftly in a decade. Especially, Consumer Generated Media (CGM) services such as Weblog and Social Networking Service (SNS) have been getting more familiar to general internet users. As users send messages actively on the web, the number of cybercrimes such as a phishing has increased. Information security services and technologies like WebTrust for CA (Certification Authority) are provided to take measures against it. However all users don’t necessarily know the meaning of these information security services and technologies (Dhamija et al., 2006). Therefore "information credibility study" became the main controversial themes in the field of Human –Computer Interaction.

The first study on the credibility of web sites was done by Fogg et al. (Fogg et al., 2001a; 2001b; 2002; 2003). Through questionnaires and experiments, they showed that the factors that effected a user's evaluation of the credibility of a site most strongly were the design and the information provided. The authors have also adopted an approach similar to Fogg et al., deriving a method which evaluates website credibility from elements including design and information provided (Watanabe et al., 2007). However, evaluating based only on design and content is not able to identify ill-intentioned sites posing as trustworthy sites, such as phishing sites. This research focuses on the verifiability of the information used to evaluate trustworthiness, and studies ways of evaluating the trustworthiness of a website more accurately by using third-party information about the site information provider.

2 USER BEHAVIORS IN EVALUATING CREDIBILITY

The authors studied specific credibility evaluation models used on e-Commerce (EC) and Non-Profit Organization (NPO) websites, based on questionnaires and the results of experiments using real subjects (Watanabe et al., 2007). We observed the behaviour of subjects when evaluating credibility and explain it in terms of a three-step process.

1) Check consistency with the basic message pattern

First the website is examined and evaluated as to whether the basic information that one would naturally assume to be provided (the "basic message pattern") is actually present in an appropriate form. If there is inconsistency with the pattern the credibility evaluation drops considerably.
2) Evaluate production elements
In addition to the basic message, elements of the website such as the design, usability and detailed information are evaluated. If the elements are appropriate for the website, the site is deemed more trustworthy, but if there are inadequacies, the evaluation suffers.

3) Check information reliability from outside the website
The evaluation in 1) and 2) is based on information found in the site itself. Users also look for and check third-party information outside of the site in question to confirm and complement this evaluation. External information tends to be consulted especially if the evaluation in 1) and 2) is not particularly good, but all users do not always perform this step.

Of the three steps above, the authors have derived an evaluation model corresponding to 1) and 2) earlier, but did not consider the process of checking external indicators as in 3). The research due to Fogg et al. also did not consider the factors in step 3).

The information on a website consists of the site's own statements about itself, so actually, a verifiably correct evaluation based only on this information is not possible. In real society, we also do not evaluate credibility based on a person's own statements about themselves, but generally seek third-party information to verify it.

However, in the experiments we have conducted so far, there are in fact very few users that actually use external information to check the credibility of a website.

3 CREDIBILITY EVALUATION FRAMEWORK
It is difficult to identify ill-intentioned websites, such as phishing sites, when evaluating the credibility of the site based only on clues in the site itself.

In experiments in which subjects were shown a website and asked to determine whether it was a phishing site (Dhamija et al., 2006), 23% of the subjects only looked at the site content, and did not check other factors like the contents of the address bar or whether the SSL-lock icon was displayed. Most of the subjects did not understand the meaning of the SSL warning messages, and they reported that, indeed, elaborate phishing sites with well-designed logos and icons were able to fool 90% of the subjects.

In other words, there are limitations to users' ability to recognize phishing sites, and those with a design that simply looks trustworthy may often be successful.

The experiments also showed that no matter how much the content of the site is analyzed, it will not be possible to accurately evaluate the reliability of the site.

The goal of our research is not to identify phishing sites, but to more-accurately evaluate the credibility of websites, and study schemes to support the users' ability to make this judgment. In particular, we expect to be able to support users with particularly low Internet literacy in this way.

Towards this goal, the authors considered the following three approaches to determining the credibility of a website.

1) Evaluate clues internal to the website
This is also done by Watanabe et al. (2007). The first thing the user sees is the website itself, so evaluating it is essential to evaluating the credibility of the site.

2) Evaluation of the information provider based on third-party information
Examining the details of information provided on third-party websites referencing the site in question should be helpful in evaluating the credibility of the site. As mentioned earlier, however, the number of users checking third-party information is not particularly high.

As such, it should be helpful, particularly for users with low information literacy, if the system can perform this type of evaluation and display the results to the user on a regular basis.

3) Evaluation based on hyperlink structures
Phishing sites often use the names of reliable information providers while carrying on fraudulent behavior, so it is difficult to correctly evaluate credibility based on name alone.

According to a survey by the Anti-Phishing Working Group (APWG) in the USA, the average amount of time a phishing site exists is very short; about four days (APWG, 2007), so it is not likely that there will be any links from other sites to the site. It may be possible to evaluate the credibility of a site by analyzing the structure of hyperlinks to the site.

As mentioned in the definitions earlier, credibility is something that the user him/herself must decide. There is a need to support better decision making about credibility, and evaluating
websites based on the above three points and having the result displayed for the user should help even low-information-literacy users make appropriate decisions about the credibility of websites.

In this paper, we discuss evaluation based on third-party information about the information provider in particular detail.

4 WAYS TO EVALUATE CREDIBILITY OF THE INFORMATION PROVIDER

User reviews have become a widely-used approach to evaluating the credibility of websites using third-party information. Some EC websites incorporate comments and ratings from users that have already made purchases through the site. Displaying evaluations from existing customers is an indirect way of expressing the credibility of the product or site operator itself, but there are still the problems of whether the provider of the comment or the comment itself is actually reliable.

As an example, one could search the web for the name of the business in order to gather third-party information about the operator of an EC website, but it is still difficult to determine which of the results are reliable.

So, the authors focused on any public activities of the business or organization acting as the information provider. In other words, we looked at references to the name of the organization in information published on the websites of organizations that are more public in nature.

4.1 Validity of the Evaluation Method

For this study, we targeted businesses operating e-Commerce (EC) websites. EC sites will have been legally required to register or apply for various permits and licenses, file reports, and have a history of affiliation with public institutions and business associations. Businesses working with public institutions are also often required to go through an investigation process. At minimum, insubstantial companies and organizations are not likely to be able to work with public institutions. Much of this sort of public activity is recorded on the websites of public institutions such as governments, municipalities and other administrative organizations.

Other information like certification levels (ISO9000, ISO14000, Privacy Certification, etc.) can also be used. These are systems for certifying organizational activities, so they can also be a source of information to verify the credibility of the information provider.

4.2 Preliminary Research

EC sites are not necessarily operated by major companies, and in fact, many are operated by small and medium-sized businesses. Because of this, there is some doubt about whether a given company’s name will be referenced on government or municipal websites, so we performed a survey using a sample of real companies.

4.2.1 Method

We first selected an arbitrary 246 Japanese EC websites, being sure to also include sites that are less well-known.

We then used a search engine (Google), searching for the company’s official name to examine the amount of information available on government sites (go.jp domain), regional municipality sites (le.jp, pref.*,jp, etc.), and websites of public organizations (or.jp).

Then we looked at up to 100 results more closely, and classified them according to type of reference.

Note that domain names in these domains can only be obtained by organizations of certain types, and organizations must provide documentation that they qualify for the domain name.

4.2.2 Results

This investigation is still in progress, so the results below represent only a partial survey.

Overall, 226 of the 246 companies (91.9%) were referenced on other websites with public domain names, which is relatively high (see Table 1).

Table 1: Rate of reference in public domain sites.

<table>
<thead>
<tr>
<th>Domain</th>
<th>go.jp</th>
<th>or.jp</th>
<th>lg.jp</th>
<th>pref.*.jp</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rate of appearance (%)</td>
<td>76.4%</td>
<td>89.4%</td>
<td>46.7%</td>
<td>67.5%</td>
</tr>
<tr>
<td>Avg. no. of references</td>
<td>598.3</td>
<td>372.3</td>
<td>18.4</td>
<td>73.9</td>
</tr>
</tbody>
</table>

(N=246)

The content of the references were classified into the 15 categories below (Table 2).

As can be seen from this classification, even in references in websites with public domains there are items related to credibility and others that are not. If
the evaluation can be done in consideration of this type of difference in the references it should be possible to further improve the accuracy.

Table 2: Avg. number of appearances per company.

<table>
<thead>
<tr>
<th>Type</th>
<th>Number</th>
<th>Type</th>
<th>Number</th>
</tr>
</thead>
<tbody>
<tr>
<td>Permit</td>
<td>0.57</td>
<td>Organization introduction</td>
<td>20.20</td>
</tr>
<tr>
<td>Registration</td>
<td>4.21</td>
<td>Committee</td>
<td>11.32</td>
</tr>
<tr>
<td>Certification/Award</td>
<td>3.92</td>
<td>Financial Reporting</td>
<td>2.67</td>
</tr>
<tr>
<td>Member of a public agency</td>
<td>4.70</td>
<td>Bankruptcy/ Litigation</td>
<td>0.25</td>
</tr>
<tr>
<td>Bidding/Contracts</td>
<td>5.74</td>
<td>Recalls, etc.</td>
<td>3.81</td>
</tr>
<tr>
<td>Delivery/Provisioning</td>
<td>1.09</td>
<td>others</td>
<td>6.11</td>
</tr>
<tr>
<td>Participation in public activities</td>
<td>14.60</td>
<td>outside object</td>
<td>22.76</td>
</tr>
</tbody>
</table>

(N=246)

4.2.3 Challenges

This method also has limitations. One limitation is that the name of the organization may not necessarily be unique in Japan, and it may be difficult to distinguish between organizations in these cases. In this study, we were able to reduce the amount of this sort of confusion by using the official name of the organization, but it will be necessary to study more-accurate and effective ways to resolve this difficulty.

5 DISCUSSION

In this paper we have proposed a method for evaluating the credibility of websites that uses third-party information to verify the credibility of the site's information provider in addition to the site design and the information provided on the website itself.

As discussed in the definitions section, credibility is something that users must decide for themselves, but as shown in this paper, we believe that gathering appropriate third-party information can help users make this sort of determination more accurately.

More specifically, the user's evaluation of the credibility of a website can be verified using third-party information. If the evaluation is correct, the user's confidence can be raised by the amount and quality of the information from third-party sources. In other words, if there is very little information available from third-party sites, there is more risk regarding whether the user's evaluation is correct or not.

From this perspective, integrating third-party information provides another indicator to support decisions about credibility for low-information-literacy users.

6 FUTURE WORK

In the future, we plan to develop a system which implements the three approaches described here and to evaluate the effectiveness of the methods.

REFERENCES


<table>
<thead>
<tr>
<th>Author Name</th>
<th>Page Numbers</th>
</tr>
</thead>
<tbody>
<tr>
<td>Abrahamsson, P</td>
<td>169</td>
</tr>
<tr>
<td>Abreu, P</td>
<td>51</td>
</tr>
<tr>
<td>Adebanjo, D.</td>
<td>93</td>
</tr>
<tr>
<td>Aguilera, A</td>
<td>58</td>
</tr>
<tr>
<td>Akioyoshi, M</td>
<td>447</td>
</tr>
<tr>
<td>Alahuta, P</td>
<td>169</td>
</tr>
<tr>
<td>Al-Bayari, O.</td>
<td>217</td>
</tr>
<tr>
<td>Alboae, L.</td>
<td>64</td>
</tr>
<tr>
<td>Ali, H.</td>
<td>209</td>
</tr>
<tr>
<td>Ali, S.</td>
<td>209</td>
</tr>
<tr>
<td>Al-Omar, K</td>
<td>191</td>
</tr>
<tr>
<td>Alotaibi, M</td>
<td>143</td>
</tr>
<tr>
<td>Ando, M.</td>
<td>484</td>
</tr>
<tr>
<td>Andonoff, E.</td>
<td>267</td>
</tr>
<tr>
<td>Arba, R.</td>
<td>437</td>
</tr>
<tr>
<td>Arsic, M.</td>
<td>203</td>
</tr>
<tr>
<td>Aryanto, V.</td>
<td>464</td>
</tr>
<tr>
<td>Asensio, J.</td>
<td>259</td>
</tr>
<tr>
<td>Ayala, R.</td>
<td>259</td>
</tr>
<tr>
<td>Baek, E.</td>
<td>241</td>
</tr>
<tr>
<td>Baptista, F.</td>
<td>225</td>
</tr>
<tr>
<td>Bara, A.</td>
<td>460</td>
</tr>
<tr>
<td>Barry, C.</td>
<td>339</td>
</tr>
<tr>
<td>Bernardos, A.</td>
<td>177</td>
</tr>
<tr>
<td>Biffl, S.</td>
<td>454</td>
</tr>
<tr>
<td>Bloice, M.</td>
<td>235</td>
</tr>
<tr>
<td>Bologa, A.</td>
<td>425, 460</td>
</tr>
<tr>
<td>Bologa, R.</td>
<td>425, 460</td>
</tr>
<tr>
<td>Bouaziz, R.</td>
<td>267</td>
</tr>
<tr>
<td>Bouzguenda, L.</td>
<td>267</td>
</tr>
<tr>
<td>Braghetto, K.</td>
<td>323</td>
</tr>
<tr>
<td>Brambring, T.</td>
<td>155</td>
</tr>
<tr>
<td>Bruno, G.</td>
<td>315</td>
</tr>
<tr>
<td>Buhl, M.</td>
<td>285</td>
</tr>
<tr>
<td>Bulander, R.</td>
<td>413</td>
</tr>
<tr>
<td>Bulo, I.</td>
<td>37</td>
</tr>
<tr>
<td>Butler, S.</td>
<td>93</td>
</tr>
<tr>
<td>Casar, J.</td>
<td>177</td>
</tr>
<tr>
<td>Chaâbane, M.</td>
<td>267</td>
</tr>
<tr>
<td>Charletton, D.</td>
<td>339</td>
</tr>
<tr>
<td>Chen, D.</td>
<td>433</td>
</tr>
<tr>
<td>Cho, V.</td>
<td>279, 421</td>
</tr>
<tr>
<td>Cornacchia, M.</td>
<td>315</td>
</tr>
<tr>
<td>Cruz, P.</td>
<td>300</td>
</tr>
<tr>
<td>Cunha, A.</td>
<td>149</td>
</tr>
<tr>
<td>Davcev, D.</td>
<td>203</td>
</tr>
<tr>
<td>Decker, M.</td>
<td>185, 255</td>
</tr>
<tr>
<td>Dominguez, A.</td>
<td>37</td>
</tr>
<tr>
<td>Durante, M.</td>
<td>37</td>
</tr>
<tr>
<td>Eckert, K.</td>
<td>307</td>
</tr>
<tr>
<td>Eriksson, N.</td>
<td>115</td>
</tr>
<tr>
<td>Ervasti, M.</td>
<td>131</td>
</tr>
<tr>
<td>Evens, T.</td>
<td>124</td>
</tr>
<tr>
<td>Fernandes, V.</td>
<td>149</td>
</tr>
<tr>
<td>Ferreira, J.</td>
<td>323</td>
</tr>
<tr>
<td>Florentina, C.</td>
<td>77</td>
</tr>
<tr>
<td>Freitas, C.</td>
<td>378</td>
</tr>
<tr>
<td>Gouveia, J.</td>
<td>161</td>
</tr>
<tr>
<td>Guido, A.</td>
<td>101</td>
</tr>
<tr>
<td>Guo, J.</td>
<td>81</td>
</tr>
<tr>
<td>Hafeez, K.</td>
<td>362</td>
</tr>
<tr>
<td>Haidegger, W.</td>
<td>107</td>
</tr>
<tr>
<td>Håkkiø, J.</td>
<td>247</td>
</tr>
<tr>
<td>Hall, J.</td>
<td>307, 390</td>
</tr>
<tr>
<td>Hattori, T.</td>
<td>29</td>
</tr>
<tr>
<td>Helaakoski, H.</td>
<td>131</td>
</tr>
<tr>
<td>Hiramatsu, A.</td>
<td>229</td>
</tr>
<tr>
<td>Höller, M.</td>
<td>235</td>
</tr>
<tr>
<td>Holzinger, A.</td>
<td>235, 454</td>
</tr>
<tr>
<td>Hu, Z.</td>
<td>81</td>
</tr>
<tr>
<td>Hung, H.</td>
<td>279</td>
</tr>
<tr>
<td>Ilievski, D.</td>
<td>203</td>
</tr>
<tr>
<td>Ioana, B.</td>
<td>77</td>
</tr>
<tr>
<td>Iribarne, L.</td>
<td>259</td>
</tr>
<tr>
<td>Ismail, H.</td>
<td>93</td>
</tr>
<tr>
<td>Iulian, P.</td>
<td>77</td>
</tr>
<tr>
<td>Iwashita, M.</td>
<td>12</td>
</tr>
<tr>
<td>Katai, O.</td>
<td>29</td>
</tr>
<tr>
<td>Kaufman, V.</td>
<td>397</td>
</tr>
<tr>
<td>Kawakami, H.</td>
<td>29</td>
</tr>
<tr>
<td>Keoy, K.</td>
<td>362</td>
</tr>
<tr>
<td>Kittl, B.</td>
<td>285</td>
</tr>
<tr>
<td>Kobayashi, Y.</td>
<td>447</td>
</tr>
<tr>
<td>Kodama, T.</td>
<td>21</td>
</tr>
<tr>
<td>Komoda, N.</td>
<td>349</td>
</tr>
<tr>
<td>Koukoulas, S.</td>
<td>390</td>
</tr>
<tr>
<td>Kunii, T.</td>
<td>21</td>
</tr>
<tr>
<td>Kuwana, T.</td>
<td>447</td>
</tr>
<tr>
<td>Lam, T.</td>
<td>476, 473</td>
</tr>
<tr>
<td>Laukkonen, T.</td>
<td>300</td>
</tr>
</tbody>
</table>
**AUTHOR INDEX (CONT.)**

<table>
<thead>
<tr>
<th>Author</th>
<th>Pages</th>
</tr>
</thead>
<tbody>
<tr>
<td>Lee, J.</td>
<td>241</td>
</tr>
<tr>
<td>Lewandowski, J.</td>
<td>285</td>
</tr>
<tr>
<td>Lieu, G.</td>
<td>421</td>
</tr>
<tr>
<td>Lima, O.</td>
<td>476</td>
</tr>
<tr>
<td>Lin, C.</td>
<td>384, 443</td>
</tr>
<tr>
<td>Liu, W.</td>
<td>433</td>
</tr>
<tr>
<td>Livi, S.</td>
<td>315</td>
</tr>
<tr>
<td>Lukovic, S.</td>
<td>197</td>
</tr>
<tr>
<td>Määttä, J.</td>
<td>247</td>
</tr>
<tr>
<td>Mahfuz, S.</td>
<td>480</td>
</tr>
<tr>
<td>Marcos, D.</td>
<td>177</td>
</tr>
<tr>
<td>Marez, L.</td>
<td>124, 392</td>
</tr>
<tr>
<td>Marin, P.</td>
<td>77</td>
</tr>
<tr>
<td>Martins, M.</td>
<td>370</td>
</tr>
<tr>
<td>Mendes, L.</td>
<td>89, 476</td>
</tr>
<tr>
<td>Mendes, P.</td>
<td>51</td>
</tr>
<tr>
<td>Meroño-Cerdan, A.</td>
<td>405</td>
</tr>
<tr>
<td>Moritsu, T.</td>
<td>349</td>
</tr>
<tr>
<td>Muntean, M.</td>
<td>425</td>
</tr>
<tr>
<td>Muresan, L.</td>
<td>437</td>
</tr>
<tr>
<td>Natchetoi, Y.</td>
<td>397</td>
</tr>
<tr>
<td>Nicolò, E.</td>
<td>315</td>
</tr>
<tr>
<td>Nishimatsu, K.</td>
<td>12</td>
</tr>
<tr>
<td>Nose, K.</td>
<td>229</td>
</tr>
<tr>
<td>Nummiaho, A.</td>
<td>169</td>
</tr>
<tr>
<td>Oliveira, J.</td>
<td>378</td>
</tr>
<tr>
<td>Oliveira, P.</td>
<td>161</td>
</tr>
<tr>
<td>Oliveira, T.</td>
<td>370</td>
</tr>
<tr>
<td>Padilla, N.</td>
<td>259</td>
</tr>
<tr>
<td>Pah, I.</td>
<td>45</td>
</tr>
<tr>
<td>Paiano, R.</td>
<td>101</td>
</tr>
<tr>
<td>Panhan, A.</td>
<td>89, 476</td>
</tr>
<tr>
<td>Papa, F.</td>
<td>315</td>
</tr>
<tr>
<td>Pattinson, C.</td>
<td>5, 73</td>
</tr>
<tr>
<td>Podean, M.</td>
<td>437</td>
</tr>
<tr>
<td>Ponomarev, V.</td>
<td>397</td>
</tr>
<tr>
<td>Popa, E.</td>
<td>45</td>
</tr>
<tr>
<td>Pu, C.</td>
<td>323</td>
</tr>
<tr>
<td>Pulimeno, E.</td>
<td>101</td>
</tr>
<tr>
<td>Puzovic, N.</td>
<td>197</td>
</tr>
<tr>
<td>Rigas, D.</td>
<td>143, 191</td>
</tr>
<tr>
<td>Robinson, L.</td>
<td>73</td>
</tr>
<tr>
<td>Sabau, G.</td>
<td>425</td>
</tr>
<tr>
<td>Sadoun, B.</td>
<td>217</td>
</tr>
<tr>
<td>Santos, D.</td>
<td>89</td>
</tr>
<tr>
<td>Sapio, B.</td>
<td>315</td>
</tr>
<tr>
<td>Saraiva, J.</td>
<td>225</td>
</tr>
<tr>
<td>Schiefer, G.</td>
<td>255</td>
</tr>
<tr>
<td>Schuurman, D.</td>
<td>124</td>
</tr>
<tr>
<td>Seki, Y.</td>
<td>21</td>
</tr>
<tr>
<td>Shimogawa, S.</td>
<td>12</td>
</tr>
<tr>
<td>Shiose, T.</td>
<td>29</td>
</tr>
<tr>
<td>Siddiqui, T.</td>
<td>5</td>
</tr>
<tr>
<td>Silva, A.</td>
<td>225</td>
</tr>
<tr>
<td>Sonehara, N.</td>
<td>484</td>
</tr>
<tr>
<td>Soto-Acosta, P.</td>
<td>405</td>
</tr>
<tr>
<td>Stanisavljevic, M.</td>
<td>197</td>
</tr>
<tr>
<td>Stoica, F.</td>
<td>45</td>
</tr>
<tr>
<td>Strandvik, P.</td>
<td>115</td>
</tr>
<tr>
<td>Strangio, M.</td>
<td>334</td>
</tr>
<tr>
<td>Su, Y.</td>
<td>443</td>
</tr>
<tr>
<td>Subero, A.</td>
<td>58</td>
</tr>
<tr>
<td>Sung, S.</td>
<td>241</td>
</tr>
<tr>
<td>Susaki, S.</td>
<td>429</td>
</tr>
<tr>
<td>Takahashi, Y.</td>
<td>429</td>
</tr>
<tr>
<td>Taniguchi, Y.</td>
<td>447</td>
</tr>
<tr>
<td>Thissen, D.</td>
<td>155</td>
</tr>
<tr>
<td>Tili, M.</td>
<td>476</td>
</tr>
<tr>
<td>Torabi, T.</td>
<td>209</td>
</tr>
<tr>
<td>Trauner, J.</td>
<td>454</td>
</tr>
<tr>
<td>Urlesberger, B.</td>
<td>235</td>
</tr>
<tr>
<td>Usugami, J.</td>
<td>357</td>
</tr>
<tr>
<td>Varajão, J.</td>
<td>149, 161</td>
</tr>
<tr>
<td>Verdegem, P.</td>
<td>292</td>
</tr>
<tr>
<td>Vinhas, V.</td>
<td>51</td>
</tr>
<tr>
<td>Wallin, A.</td>
<td>247</td>
</tr>
<tr>
<td>Watanabe, K.</td>
<td>484</td>
</tr>
<tr>
<td>Wong, Y.</td>
<td>279</td>
</tr>
<tr>
<td>Yamada, T.</td>
<td>429</td>
</tr>
<tr>
<td>Yamasaki, T.</td>
<td>229</td>
</tr>
<tr>
<td>Yamasaki, T.</td>
<td>362</td>
</tr>
<tr>
<td>Zhu, D.</td>
<td>384, 443</td>
</tr>
</tbody>
</table>