# MUMIS -- ADVANCED INFORMATION EXTRACTION FOR MULTIMEDIA INDEXING AND SEARCHING[*]

T. DECLERCK

*DFKI GmbH, Language Technology Department*
*Stuhlsatzenhausweg 3,*
*D-66123 Saarbruecken, Germany*
*E-mail: declerck@dfki.de*

H. CUNNINGHAM and H. SAGGION

*Department of Computer Science, University Sheffield*
*Regent Court, 211 Portobello St., Sheffield S1 4DP, UK*
*E-mail: {Hamish|H.Saggion}@dcs.shef.ac.uk*

J. KUPER and D. REIDSMA

*Department of Computer Science, University of Twente*
*PO Box 217, 7500 AE Enschede, The Netherlands*
*E-mail: {jankuper|dennisr}@cs.utwente.nl*

P. WITTENBURG

*Max Planck Institute for Psycholinguistics*
*Wundtlaan 1, PB 310, 6500 AH Nijmegen, The Netherlands*
*E-mail: Peter.Wittenburg@mpi.nl*

This paper describes the role advanced *natural language processing* (NLP) and especially *information extraction* (IE) can play for multimedia applications. As an example of such an application, we present an approach dealing with the automatic conceptual indexing of multimedia documents, which subsequently can be searched by semantic categories instead of key words. A novelty of the approach is to exploit multiple sources of information relating to video content. In the MUMIS scenario, the source of information consists in a rich range of textual and transcribed sources covering soccer games.

## 1. Introduction

Multimedia repositories of moving images, texts, and speech are becoming increasingly available. This together with the needs for 'video-on-demand' systems require fine-grain indexing and retrieval mechanisms allowing users access to specific segments of the repositories containing specific types of information. Annotation of video is usually carried out by humans that follow strict guidelines, which foresee the annotation with 'metadata' such as people

[*] This work has been supported by EC grant IST-1999-10651 for the MUMIS project.

involved in the production of the visual record, places, dates, and keywords that capture the essential content of what is depicted. Still, there are a few problems with human annotation: 1) The cost and time involved in the production of 'surrogates' of the programme is very high; and 2) Humans are subjective when assigning descriptions to visual records; and 3) the level of annotation required to satisfy user's need can hardly be achieved with the use of mere keywords.

In order to tackle these problems, content-based (or visually-based) methods have risen (see [13]). Content-based indexing and retrieval of visual records is based on features such as colour, texture, and shape. Yet visual understanding is not well advanced and is very difficult even in closed domains. For example, visual analysis of the video of a football match can lead to the identification of interesting "content" like a shooting scene (i.e., the ball moving towards the goal) [2, 7], but this image analysis approach will hardly ever detect who is the main actor involved in that scene (i.e., the shooter). As a consequence, many research projects have explored the use of linguistic analysis of collateral textual descriptions of the images (either still or moving) for automatic tasks such as indexing [6, 11], classifying [9], or understanding [11, 12] of visual records.

## 2. MUMIS: a Multimedia Indexing and Searching Environment

MUMIS is proposing an integrated solution to the problem of multimedia indexing and searching. The solution consists in applying advanced *natural language processing* (NLP) on different sources (structured, semi-structured, free, etc.), modalities (text, speech), and languages (English, German, Dutch) all describing the same event to carry out database population, indexing, and search. For this purpose the project makes also use of domain ontologies and of a specialized set of lexicons for the selected domain (soccer). MUMIS thus makes intensive use of the resulting linguistic and semantic based annotations (see [3] for more details on linguistic and semantic annotations), coupled with domain-specific information, in order to generate formal annotations of events that can serve as index for videos querying (see also [5, 14] for more details).

The core linguistic processing for the annotation of the multimedia material consists of *information extraction* (IE) techniques for identifying, collecting and normalizing significant text elements (such as the names of players in a team, goals scored, time points or sequences etc.) which are critical for the appropriate annotation of the multimedia material in the case of soccer. One system per language has been used or developed. Each system delivers an

XML output.

The novelty of the approach is not only the use of these 'heterogeneous' sources of information but also combination or cross-source fusion of the information obtained from each source. A process of alignment and rule-based reasoning that also uses the semantic model *merges* the result of all XML encoded information extraction systems. The merged annotations are then stored in a database, where they will be combined with relevant metadata that are also automatically extracted from the textual documents.

Keyframes extraction from MPEG movies around a set of pre-defined time marks - result of the information extraction component - is being carried out to populate the database. JPEG keyframes images are extracted that serve for quick inspection in the user interface. The software used for off-line keyframe extraction takes a movie file, a list of times stamps, and the size of the keyframe and produces a list of keyframes. The on-line part of MUMIS consists of a state of the art user interface allowing the user to query the multimedia database (e.g., "The corner involving Beckham"). The user is first presented with selected video key-frames as thumbnails that can be played obtaining the corresponding video and audio fragments, as can be seen in the following screen shot:

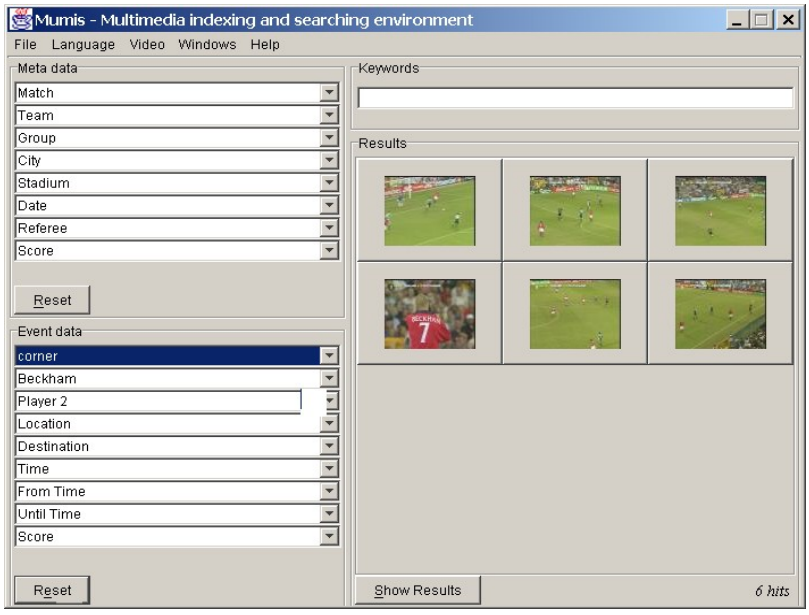


Figure 1: A screen shot of the MUMIS demonstrator: the results of the query: "Show me the corners involving Beckham". All the annotations used for indexing the video for this

kind of query have been automatically generated by the IE systems and combined in one set of merged searchable annotations.

## References


1. E. André, *Natural Language in Multimedia/Multimodal Systems*. In R. Mitkov (Ed), *Handbook of Computational Linguistics*, Oxford (2000).
2. J. Assfalg, M. Bertini, C. Colombo and A. Del Bimbo, *Semantic annotations of sports videos*. In Proceedings of the Conference on Content-Based Multimedia Indexing, CBMI-2001, Brescia (2001).
3. P. Buitelaar and T. Declerck, *Linguistic Annotation for the Semantic Web*. In S. Handschuh and S. Staab (Eds). *Annotation for the Semantic Web*. To appear (2003).
4. S.F. Chang, W. Chen, H.J. Meng, H. Sundaram and D. Zhong, *A Fully Automated Content-based Video Search Engine Supporting Spatio Temporal Queries*. IEEE Transactions on Circuits and Systems for Video Technology (1998).
5. T. Declerck, P. Wittenburg, H. Cunningham, *The Automatic Generation of Formal Annotations in a Multimedia Indexing and Searching Environment*. Proceedings of the Workshop on Human Language Technology and Knowledge Management, ACL (2001).
6. F. de Jong, J. Gauvin, D. Hiemstra, K. Netter, *Language-Based Multimedia Information Retrieval*. In Proceedings of the 6th Conference on Recherche d'Information Assistée par Ordinateur, RIAO (2000).
7. Y. Gong, L.T. Sin, C.H. Chuan, H. Zhang and M. Sakauchi, *Automatic Parsing of TV Soccer Programs.* Proceedings of the International Conference on Multimedia Computing and Systems (1995).
8. M.R. Naphade and T.S. Huang, *Recognizing high-level concepts for video indexing.* In Proceedings of the Conference on Content-Based Multimedia Indexing, CBMI, Brescia (2001).
9. C. Sable and V. Hatzivassiloglou, *Text-based approaches for the categorization of images*. Proceedings of ECDL (1999).
10. P. Salembier, *An overview of Mpeg-7 multimedia description schemes and of future visual information challenges for content-based indexing*. In Proceedings of the Conference on Content-Based Multimedia Indexing, CBMI, Brescia (2001).
11. A. Salway, *Talking Pictures: Indexing and Representing Video with Collateral Texts*. In Hiemstra D., de Jong F., Netter K. (Eds), *Language Technology in Multimedia Information Retrieval*, Enschede (1998).
12. R.K. Srihari, *Automatic Indexing and Content-Based Retrieval of Captioned Images*, Computer 28/9 (1995).
13. R. Veltkamp and M. Tanase, *Content-based Image Retrieval Systems: a survey*. Technical report UU-CS-2000-34, Utrecht University, 2000.
14. http://parlevink.cs.utwente.nl/projects/mumis/