The Right Expert at the Right Time and Place
from expertise identification to expertise selection

Pavel Serdyukov[1], Ling Feng[2], Arthur van Bunningen[3], and Sander Evers[1], Harold van Heerde[1] and Peter Apers[1], Maarten Fokkinga[1], Djoerd Hiemstra[1]

[1] Database Group, University of Twente, Enschede, The Netherlands
{serdyukovpv,everss,h.j.w.vanheerde,apers,fokkinga,hiemstra}@cs.utwente.nl
[2] Database group, Dept. of Computer Science and Tech, Tsinghua University, China
fengling@mail.tsinghua.edu.cn
[3] Teezir, The Netherlands
arthur@vanbunningen.com

Abstract. We propose a unified and complete solution for expert finding in organizations, including not only expertise *identification*, but also expertise *selection* functionality. The latter two include the use of implicit and explicit *preferences* of users on meeting each other, as well as *localization* and *planning* as important auxiliary processes. We also propose a solution for *privacy protection*, which is urgently required in view of the huge amount of privacy sensitive data involved. Various parts are elaborated elsewhere, and we look forward to a realization and usage of the proposed system as a whole.

1 Introduction

Expertise sharing is gaining increasing popularity and importance for enterprises due to the fact that a mass of knowledge has been accumulated in the course of corporate business, and meanwhile employees tend to seek and interact with knowledgeable people for information prior to using some formal sources to solve their daily work related problems. It is even common that users often search for persons rather than for relevant documents [15]. Besides being sources of unpublished knowledge, the experts on the search topic are also able to explain problems and solutions by guiding the user through existing artifacts. However, attempts to identify experts by manual browsing through organizational documents or via informal social connections are impractical in large enterprises, especially when they are geographically distributed. Usually, a specialized *expert finding system* (also known as expert search, expert recommendation or expertise location system) is developed to assist in the search for individuals or departments that possess certain knowledge and skills within the enterprise and outside [26]. It allows either to save time and money on hiring a consultant when company's own human resources are sufficient, or helps to find an expert at affordable cost and convenient location in another organization.

Finding an expert is a challenging task, because expertise is a loosely defined and not a formalized notion. It is common to refer to expertise as to "tacit knowledge" [8], the type of knowledge that people carry in their minds and is, therefore, difficult to access. It is opposed to "explicit knowledge" which is easy to capture, describe, document and store. Thus, an expert finding system aims to manage "tacit knowledge" in organizations by inferring it using organizational "explicit knowledge" and finally to transfer it among people by helping their socialization and knowledge exchange. With respect to these missions, it is common to divide the task of expert finding into two stages of equal importance: *expertise identification* and *expertise selection* [27].

At the *expertise identification* stage, all employees of the organization are ranked with respect to the user information need for expertise, usually expressed in *a short text query*. It is often unclear what amount of personal knowledge should be considered enough to name somebody "an expert". It depends not only on the specificity of the user query, but also on characteristics of respective expertise area: on its age, depth and complexity. However, expert finding systems do not actually infer the level of expertise or any quantitative estimate which is easy to semantically interpret or map to a qualitative scale. They just provide some estimate that may be used to rank people by their expertise level.

Since arranging a meeting with even a single expert might be very time-consuming, a practically usable expert finder should help not only to identify knowledgeable people, but also to *select those experts* that are most appropriate for a face-to-face contact with the user [1]. Since expert finding is a tool for improving organizational communication, it must be able to predict various features of a planned communication in order to help it be successful. In the first place, it should aim for a communication that is physically doable. So, the availability and interruptability of experts that may depend on their location and/or occupancy should be considered. In other cases, an intelligent meeting planning, taking into account agenda records of several employees, including the expertise seeker, as well as predictions for their future location, is required. An expert finder should also try to predict whether the communication is likely to be desired by both parts. Various human factors like expert's mood or mental stress may be considered. Preferences of users on communication with certain people (e.g., based on their positions/ranks or reputation in a company) should also be integrated.

Introducing expert finding in an enterprise inevitably results in an environment in which highly precise data about the whereabouts of employees, their behavior, preferences and the way they spent their time will be collected. Although any goal of increasing work efficiency might be of interest for both employers and employees, and employers do have the right to monitor and collect data from their employees, such goal can only be achieved when the right of privacy is not violated in the process [28]. Unlimited and unrestricted collecting of the private data, for example by monitoring the browsing behavior of employees, will infringe the privacy of the employee, with undesirable effects. Hence, when monitoring and collecting data is inevitable in order to enable services like expert finding, a clear corporate privacy policy is needed.

Our Contribution. The expertise identification task developed a lot during last years as a subject of research on Information Retrieval [14]. Recently, we proposed several solutions that were successfully evaluated within widely accepted experimental frameworks [31–34]. Moreover, the quality of existing research prototypes is currently quite high with an error rate about 50% [35], what motivates to expand the scope of research in expert finding to a broader spectrum of related vital issues. Consequently, this paper mainly seeks to *research on expertise selection.* To the best of our knowledge, this integral stage of expert finding is traditionally neglected in contemporary approaches and practically no academic research has been conducted in this direction, even at the conceptual level. We propose a unified and complete solution for expert finding in organizations, including not only *expertise identification*, but also *expertise selection* functionality. We present methods that solve several identified problems and have recently shown themselves to advantage in the respective problem domains; these methods will hopefully ease the development of a fully fledged expertise sharing system.

Organization of the Paper. The next section gives a brief overview of existing research in expert finding. The follow-up sections propose new solutions for each of the problems that a real-world expert finding system with a complete functionality inevitably faces. Section 3 describes how to make a first step from expertise identification to expertise selection and consider a model of user preferences on meeting certain people in the organization. Section 4.1 continues to explain how to efficiently facilitate expertise selection and describes methods for measuring and monitoring experts' availability (i.e., localization) considering that all users have the implicit preference that their question is answered immediately. Section 4.2 shows how to integrate another implicit desire of most users — to meet experts eventually in the future if the question is left unanswered. Section 5 proposes a unified method for ensuring privacy based on the fact that the proposed additional expertise selection stage needs a lot of private user data for the analysis. Section 6 concludes the paper with a summary of our contribution.

2 Existing solutions

Expertise identification. In early expert finding systems the prediction of personal expertise was often made through the analysis of textual content of employee profiles. These profiles contained summaries of personal knowledge and skills, affiliations, education and interests, as well as contact information [16, 17, 9]. However, such profiles are always known to be incomplete and outdated due to serious time investment needed for their maintenance [10]. Therefore, the majority of successor systems, numerously emerged in academia during recent years, regarded any documents the person is related to as possible indicators of his/her expertise. They commonly assumed that the more often a person is related in the documents containing many words describing the topic, the more likely we may rely on such a person as on an expert.

Existing approaches to expertise identification naturally fall into two categories. *Profile-centric* approaches [23, 29] merge all documents related to a candidate expert into a single personal profile either prior to retrieval time, or dynamically using only the top retrieved documents [31]. These personal profiles are then regarded as single documents to be ranked using standard measures of document relevance. *Document-centric* approaches first rank documents and then sum their retrieval scores for each related candidate to estimate the degree of candidate's expertise [25, 7, 19]. It was also proposed to calculate only the relevance score of the text window surrounding the person's mentioning [24] and to propagate relevance from documents to their related candidates not in one-step, but in several steps through utilizing graph structure of the respective expertise domain [34].

Expertise selection. Expertise identification methods mostly develop due to the interest in the academic world; in contrast, expertise selection research is making its marginal progress only due to the existence of industrial expert finding solutions. However, even their assistance in expertise selection is not all-embracing. Some of these systems offer powerful ways to represent and manually navigate search results, what, to a certain extent, simplifies expertise selection. Autonomy (`autonomy.com`), the undoubted market leader, allows the classification of experts in the result list by competency areas and positions in a company. So does the Endeca (`endeca.com`), the third enterprise search market leader after FAST (`fastsearch.com`). In some cases, the searcher's context is not totally ignored and implicit preferences of the user on types of people are considered: Microsoft's Knowledge Network recommends those experts who are found in proximity of the user in organizational social network. Workload aspects are considered by AskMe (`askmecorp.com`) that develops an expert finder on top of the FAST platform: it enables experts to personally control or change the number of questions that they are willing to answer at any given time.

Apparently, neither academic, nor industrial approaches to expert finding are ready to facilitate expertise selection at a full-scale level. While expert finders offered on the market are of a great help to improve organizational communication and knowledge flow, they are still too far from providing a complete solution. Such a long-awaited software that would assist at each step of expertise sharing and acquisition is envisioned in early research on expert finding [27, 20], although no real solutions for design and implementation are proposed so far. Our work is the first attempt to not only decompose the expertise selection problem, but also to propose specific ways to overcome each discovered issue.

3 Expertise Selection with Explicit Preferences

While a lot of user preferences on meeting certain people are easy to infer just using common sense assumptions or global statistics of the system's usage, it is still reasonable to start the design of expertise selection component from making up a mechanism for setting up explicit user preferences on persons to communicate with. Such preferences could be of a great help for both sides: the expertise seeker and the expert who is ready to share the expertise under certain conditions. Although preferences on certain individuals are easily imagined, typically, users do not know everyone in their enterprise and hence should bid their preference only on features they like or dislike in people. In this connection, it is important to draw a line between two types of a person's features: static features whose value do not change or change slowly over time (e.g., age, gender, position in a company, education, etc.) and dynamic, or context-specific, features that may vary even within a minute (e.g., location, emotional state, workload).

In the data management field, two approaches cab be distinguished to deal with users' preferences, namely, quantitative and qualitative [13]. The qualitative approach intends to directly specify preferences between the data tuples in the query answer, typically using binary preference relations. An example preference relation is "*prefer expert A to another expert B if and only if A's rank in company X is higher than B's and A holds no part-time positions in other companies*". These kinds of preference relations can be embedded into relational query languages through relational operators or special preference constructors, which select from its input the set of the most preferred tuples (e.g., Winnow [13], PreferenceSQL [21], Skyline [12]). The quantitative approach expresses preferences using a scoring function, which associates a numeric score with every tuple of the query. A framework for expressing and combining such kinds of preferences functions was provided by Agrawal [3] and Koutrika [22]. The latter approach seems more appropriate for the expertise selection task, as well as for any task with a majority of non-binary or uncertain features.

Our design of a knowledge-based context-aware preference model allows to take set of ranked experts and *select the topmost preferred* of them by *re-ranking* the set. We use a variant of Description Logics [6] to represent preferences and then apply a probabilistic inference mechanism. Let F be a function that for each expert e gives its features: $F(e) = \{f_1, f_2, \ldots, f_n\}$. We use $Prob(f \in F(e))$ to express the probability that feature f holds for expert e. We assume that the features are independent. Similarly, let function G give for each preference p its context features (e.g., $hasStatus.\{Free\}$): $G(p) = \{g_1, g_2, \ldots, g_m\}$. (Since the context may feature different properties over time, whereas experts have features that are relatively stable, we call the latter *static* and the former *dynamic*.) The satisfaction of a context feature usually depends on measurements returned from error-prone (hence *uncertain*) sensors; therefore we use $Prob(g \in G(p))$ to denote the probability that preference p has context feature g. To decide whether a given expert satisfies the user's preference, we need to determine the probability that the expert is ideal for the required context features of the preference. To this end, let σ be the *score function* that for each pair (g, f) in the observed history H returns a score $\sigma(g, f)$: the probability in H that for a preference with context feature g, the user has selected an expert with static feature f. Now, in terms of these concepts, the probability that an expert e is the ideal one according to preference p, can be expressed. Details of our inference mechanism are described elsewhere [37].

4 Expertise Selection with Implicit Preferences

Despite the fact that explicit preferences are usually indispensable for personalized systems, users are often not enough enthusiastic to accurately specify their preference models, although still expecting the system to be efficient. However, some preferences are likely to be assumed by all users by default. When a group of two or more users agrees to meet for expertise sharing, their usual demand from the system which is responsible for arranging such a meeting is to organize it as soon as possible, considering current and future locations. If that is not possible, the next preference is to have such a meeting in any reasonable time in the future, during some common free time slot of all involved users.

4.1 Implicit preference on immediate meetings

Most requests coming from users to experts are short questions awaiting for short answers. In such cases, when a momentary communication is sufficient for the desired knowledge exchange, an expertise selection component needs just to infer current user locations, make a guess about the time all users need to approach each other and, if an immediate meeting is possible, inform all interested persons about such an opportunity. While the inference of a current user activity and his/her level of occupancy is preferable, the location context is usually selective enough to filter out a lot of opportunities. The proximity of the users to communication facilities, e.g., a videoconference system or a phone, could be also considered.

The most important information for the localization of users is usually taken from several types of "sensors". An important observation is that a lot of location information *is already present, but not exploited* in many enterprises, and our goal is to make full use of such information. Possible sources for current location information are:

- WiFi access points that register the proximity of an expert's WiFi-enabled mobile devices such as laptop or PDA. Enterprise buildings are usually well-covered by WiFi, but an expert may not always carry his/her mobile device.
- Bluetooth access points that register the proximity of cellphones. A cellphone is more often carried on the body, but Bluetooth coverage is usually sparse. However, more and more PCs are equipped with Bluetooth, so coverage is potentially high in places where people are at work.
- Registration of access cards gives a broad indication of in which (part of a) building a person is located.
- Computer activity provides an accurate indication of where the user is, given that the location of the computer is known.
- Simple webcams or microphones can pinpoint the presence of a person. Face or speech recognition may even identify this person (although the accuracy of these techniques may not be high, the combination with other sensors like Bluetooth may prove useful).

Essential characteristics for this kind of "sensor data" are that it is uncertain, incomplete and heterogeneous. The user locations returned by sensor networks are therefore always probability distributions and hence to deal with this, we use probabilistic models. They model where a person can go, what devices he has with him, and where people or devices can be sensed by several technologies. The *observed variables* consist of the signal strengths of WiFi scans, detections of Bluetooth devices, recorded computer activity, etc. Some of these observations can be modeled as *instantaneous*, others (like Bluetooth or WiFi scans) take a certain *interval* to complete. A *probabilistic model* connects the observed variables to the *query variables*: in this case, these are the locations of the experts at each point in time t. Additional unobservable variables, such as whether expert A has device B with him, may also play a role in the probabilistic model.

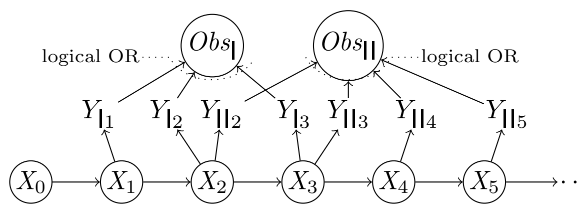

Fig. 1. Probabilistic model with Noisy-OR intervals

A general schema of our probabilistic model is shown in Fig.1. On top is a layer of observed variables; on the bottom, a layer of variables representing the *state* of the experts at each point in time. This state, represented in the figure by variable X_t, is split up into several sub-variables containing location and carried devices of each expert, but this is not shown. The graph in the figure corresponds to the structure of the Bayesian Network that defines the model; informally, an arrow pointing from X to Y means that the probability distribution of Y depends directly on the value of X. In our model, there is a layer of Y variables between states and observations. A variable Y expresses the fact that the observation of a certain device has been influenced by the location of this device at time t. This model is a variant on the Hidden Markov Model with Noisy-OR observations (HMM-NOR), and has the pleasant property that, for binary observations such as Bluetooth scans, the complexity of probabilistic inference stays linear in the length of the interval. The details of this approach are elaborated in a forthcoming paper [18].

4.2 Implicit preference on arranging meetings in the future

In cases when certain users cannot meet each other at the time of the request for expertise, their communication can be scheduled for the future with the help of an intelligent meeting planning mechanism. It is reasonable to assume that the agendas of all involved users have implicitly the user preference not to meet within occupied time slots. Considering that, the preference to meet with certain experts in the future –if the user information need will not be satisfied by that time– should not be ignored.

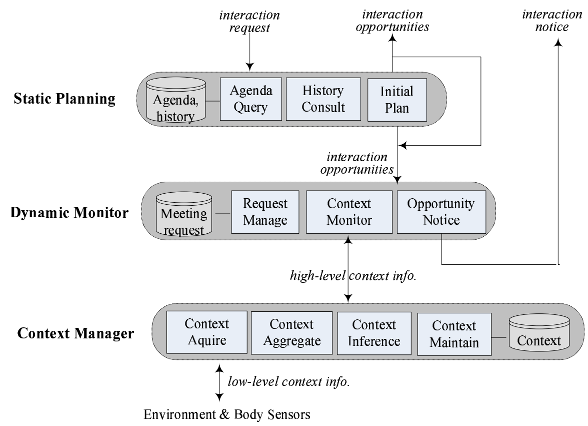

Fig. 2. Intelligent meeting planning

Figure 2 sketches how our meeting planning approach works. It first queries the agendas of both parties to find out possible time slots that satisfy requestor's constraints and preferences. Meanwhile, it also consults the previous conversation history (log) of the wanted person to enforce or complement the possibilities, which will then be returned to the user as an initial answer in the order of time. After that, it starts to monitor these possibilities, and meanwhile keeps eyes open on new unanticipated opportunities. When the behavior of the requested person deviates from the original schedule in his/her agenda, an immediate conversation might be possible as well. For instance, the requested person finishes a meeting and returns to his/her office 10 minutes earlier than scheduled, exhibiting a possibly good opportunity for a short conversation. The context monitor is responsible for checking context information in a timely fashion and decides whether a particular possibility is indeed a good conversation opportunity. Once this is the case, a chance alert will be sent to the user, so that a conversation between the user and the wanted expert can be conducted right away.

To do this, the context manager of the module plays an important role. It gathers low-level context information from various context suppliers such as sensors, and performs context aggregation and context inference so as to derive high-level context. Necessary context is stored into a context database for later retrieval and analysis. Besides, the context manager has the duty to answer pull-context queries, and actively execute push-context actions, in response to the requests of the context monitor. The final answer to the request will be logged (memorized), so that the context manager can do learning and reasoning in order to deliver smart solutions later on.

5 Privacy Control

Our approach to expert selection and meeting uses a lot of data about the people involved. In corporate environments, access control techniques [2, 11] are insufficient for privacy protection, since access control can easily be bypassed by system administrators and the employers themselves. Employees have to fully trust their employers and such trust they will not put forever, especially not in cases where there is a conflict between employer and employee. Moreover, in cases when the enterprise is a subject to investigation by governmental organizations, the stored data will be the subject of investigation too, and possibly even disclosed to the public afterwards. The Enron fraud investigation in 2002 exemplifies this [30]. Hence, although it is tempting to store all data, a balance is needed between infinite storage—keeping full potential for new or existing services—and no storage at all to make sure that privacy sensitive data will never be disclosed.

We propose a new technique termed *data degradation*, which is based on the assumption that long lasting purposes can often be fulfilled with less accurate data. Privacy sensitive data will progressively be degraded from *most accurate* via *generalized intermediate states*, up to complete removal from the system. The data degradation technique can be applied to both the data used at expertise identification stage (e.g., to the user web browsing history) and the data used at the stage of expertise selection (e.g., location traces). Using this technique, the enterprise is urged to carefully think about the form and period they need and want to store data, and gives employees the possibility to express what they find acceptable in terms of privacy, with a useful compromise between data usability and privacy as a result. The privacy benefit for employees is that they do not have to worry that the collected data about them can be misused in the future.

We consider the collected data as being a collection of *trails* of employees. A trail consists of a set of attributes, some of which are considered as privacy sensitive. In our data degradation model, termed the *life cycle policy model*, a trail is subject to progressive degradation from the accurate state to less detailed intermediate states, up to disappearance from the database. The degradation of each piece of information is captured by a *Generalization Tree*. Given a *domain generalization hierarchy* for an attribute, a generalization tree for that attribute gives, at various levels of accuracy, the values that the attribute can take during its lifetime. Hence, a path from a particular node to the root of the tree expresses all degraded forms the value of that node can take in its domain (see Figure 3). Note however, that when storing privacy sensitive data in regular databases, it must be ensured that at each degradation step, the data will be irreversibly removed from the system, which is not a straightforward task [4, 36]. For more details of the model we refer to [5].

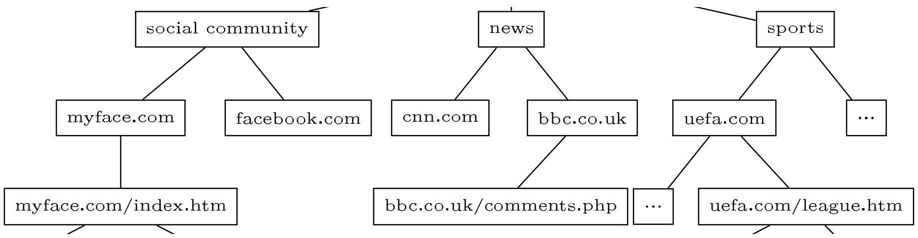

Fig. 3. Example of a generalization tree for the browsed web page attribute.

6 Conclusion

This paper reports on our joint integral efforts towards a full-fledged expertise sharing solution for enterprises, covering expertise identification and expertise selection stages. The latter includes the use of implicit and explicit preferences of users on meeting each other. We also proposed a solution for privacy protection, which is applicable at all stages of the expert finding process. The work presented here still remains at a rather preliminary stage with a number of interesting issues to be addressed in the near future. So far we have designed the system more from the functionality aspect, and have ignored the efficiency perspective. For instance, a user's expert finding task can follow different execution plans, where all involved modules can execute in different sequential orders so as to minimize expert selectivity. Besides, implementing the whole system and using it in a real setting is a necessity for the proof of concept.

Acknowledgments. This work is funded by the Dutch Organization for Scientific Research (NWO-VIDI project), Dutch Ministry of Economic Affairs (MultimediaN-AmbientDB project), and Centre for Telematics and Information Technology of the University of Twente in the Netherlands.

References

1. M. S. Ackerman, V. Wulf, and V. Pipek. *Sharing Expertise: Beyond Knowledge Management.* MIT Press, Cambridge, MA, USA, 2002.
2. R. Agrawal, J. Kiernan, R. Srikant, and Y. Xu. Hippocratic databases. In *28th Int'l Conf. on Very Large Databases (VLDB), Hong Kong*, 2002.
3. R. Agrawal and E. L. Wimmers. A framework for expressing and combining preferences. In *SIGMOD '00: Proceedings of the 2000 ACM SIGMOD international conference on Management of data*, pages 297–306, New York, NY, USA, 2000. ACM Press.
4. N. Anciaux, L. Bouganim, H. J. W. van Heerde, P. Pucheral, and P. M. G. Apers. Instantdb: Enforcing timely degradation of sensitive data. In *Proceedings of the 24th International Conference on Data Engineering.* IEEE Computer Society Press, April 2008.

5. N. L. G. Anciaux, L. Bouganim, H. J. W. van Heerde, P. Pucheral, and P. M. G. Apers. Instantdb: Enforcing timely degradation of sensitive data. In *Proceedings of the 24th International Conference on Data Engineering (ICDE2008), Cancun, Mexico*, pages 1373–1375, Los Alamitos, April 2008. IEEE Computer Society Press.
6. F. Baader, D. Calvanese, D. L. McGuinness, D. Nardi, and P. F. Patel-Schneider, editors. *The Description Logic Handbook: Theory, Implementation, and Applications*. Cambridge University Press, 2003.
7. K. Balog, T. Bogers, L. Azzopardi, M. de Rijke, and A. van den Bosch. Broad expertise retrieval in sparse data environments. In *SIGIR '07: Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval*, pages 551–558, New York, NY, USA, 2007. ACM.
8. P. Baumard. *Tacit Knowledge in Organizations*. Sage Publications, Inc., Thousand Oaks, CA, USA, 2001.
9. I. Becerra-Fernandez. Facilitating the online search of experts at NASA using expert seeker people-finder. In *PAKM'00, Third International Conference on Practical Aspects of Knowledge Management*, 2000.
10. I. Becerra-Fernandez. Searching for experts on the web: A review of contemporary expertise locator systems. *ACM Trans. Inter. Tech.*, 6(4):333–355, 2006.
11. E. Bertino, J.-W. Byun, and N. Li. Privacy-preserving database systems. *Lecture Notes in Computer Science*, 3655:178–206, 2005.
12. S. Börzsönyi, D. Kossmann, and K. Stocker. The skyline operator. In *Proceedings of the 17th International Conference on Data Engineering, April 2-6, 2001, Heidelberg, Germany*, pages 421–430. IEEE Computer Society, 2001.
13. J. Chomicki. Preference formulas in relational queries. *ACM Trans. Database Syst.*, 28(4):427–466, 2003.
14. N. Craswell, A. de Vries, and I. Soboroff. Overview of the trec-2005 enterprise track. In *Proceedings of TREC-2005*, Gaithersburg, USA, 2005.
15. N. Craswell, D. Hawking, A.-M. Vercoustre, and P. Wilkins. Panoptic expert: Searching for experts not just for documents. In *Ausweb Poster Proceedings*, Queensland, Australia, 2001.
16. T. Davenport. Knowledge Management at Microsoft. White paper. `http://www.itmweb.com/essay536.htm`, January 1997.
17. T. Davenport. Ten principles of knowledge management and four case studies. *Knowledge and Process Management*, 4(3), 1998.
18. S. Evers, M. Fokkinga, and P. M. G. Apers. Probabilistic processing of interval-valued sensor data. To appear in: Proceedings of DMSN 2008, 5th International Workshop on Data Management for Sensor Networks, To be held in conjunction with VLDB 2008, August 24, 2008, Auckland, NEW ZEALAND. `http://dmsn08.cs.umass.edu/`, 2008.
19. H. Fang and C. Zhai. Probabilistic models for expert finding. In G. Amati, C. Carpineto, and G. Romano, editors, *Advances in Information Retrieval, 29th European Conference on IR Research, ECIR 2007*, volume 4425 of *Lecture Notes in Computer Science*, pages 418–430. Springer, 2007.
20. C. Johansson, P. A. V. Hall, and M. Coquard. ẗalk to paula and peter - they are experienced ̈: the experience engine in a nutshell. In *SEKE '99: Proceedings of the 11th International Conference on Software Engineering and Knowledge Engineering, Learning Software Organizations, Methodology and Applications*, pages 171–185, London, UK, 2000. Springer-Verlag.
21. W. Kießling. Foundations of preferences in database systems. In *VLDB*, pages 311–322, 2002.

22. G. Koutrika and Y. E. Ioannidis. Personalized queries under a generalized preference model. In *Proceedings of the 21st International Conference on Data Engineering, ICDE 2005, 5-8 April 2005, Tokyo, Japan*, pages 841–852. IEEE Computer Society, 2005.
23. X. Liu, W. B. Croft, and M. Koll. Finding experts in community-based question-answering services. In *CIKM '05: Proceedings of the 14th ACM international conference on Information and knowledge management*, pages 315–316, New York, NY, USA, 2005. ACM Press.
24. W. Lu, S. Robertson, A. Macfarlane, and H. Zhao. Window-based Enterprise Expert Search. In *Proceeddings of the 15th Text REtrieval Conference (TREC 2006)*, 2006.
25. C. Macdonald and I. Ounis. Voting for candidates: adapting data fusion techniques for an expert search task. In *CIKM '06: Proceedings of the 15th ACM international conference on Information and knowledge management*, pages 387–396, 2006.
26. M. T. Maybury. Expert finding systems. Technical Report MTR06B000040, MITRE Corporation, 2006.
27. D. W. McDonald and M. S. Ackerman. Just talk to me: a field study of expertise location. In *CSCW '98: Proceedings of the 1998 ACM conference on Computer supported cooperative work*, pages 315–324, New York, NY, USA, 1998. ACM Press.
28. S. Miller and J. Weckert. Privacy, the workplace and the internet. *Journal of Business Ethics*, 28(3):255–265, 2000.
29. D. Petkova and W. B. Croft. Hierarchical language models for expert finding in enterprise corpora. In *ICTAI '06: Proceedings of the 18th IEEE International Conference on Tools with Artificial Intelligence*, pages 599–608, 2006.
30. C. S. G. Rudman and R. LLP. The enron fraud. `http://www.enronfraud.com/`, June 2008.
31. P. Serdyukov and D. Hiemstra. Modeling documents as mixtures of persons for expert finding. In C. Macdonald, I. Ounis, V. Plachouras, I. Ruthven, and R. W. White, editors, *ECIR*, volume 4956 of *Lecture Notes in Computer Science*, pages 309–320. Springer, 2008.
32. P. Serdyukov, H. Rode, and D. Hiemsta. Exploiting sequential dependencies for expert finding. In *SIGIR '08: Proceedings of the 31th annual international ACM SIGIR conference on Research and development in information retrieval*, 2008.
33. P. Serdyukov, H. Rode, and D. Hiemsta. Modeling expert finding as an absorbing random walk. In *SIGIR '08: Proceedings of the 31th annual international ACM SIGIR conference on Research and development in information retrieval*, 2008.
34. P. Serdyukov, H. Rode, and D. Hiemstra. Modeling multi-step relevance propagation for expert finding. In *CIKM '08: Proceedings of the 17th ACM conference information and knowledge management*, 2008.
35. I. Soboroff, A. de Vries, and N. Craswell. Overview of the trec-2006 enterprise track. In *Proceedings of TREC-2006*, Gaithersburg, USA, 2006.
36. P. Stahlberg, G. Miklau, and B. N. Levine. Threats to privacy in the forensic analysis of database systems. In *SIGMOD '07: Proceedings of the 2007 ACM SIGMOD international conference on Management of data*, pages 91–102, New York, NY, USA, 2007. ACM Press.
37. A. H. van Bunningen, M. M. Fokkinga, P. M. G. Apers, and L. Feng. Ranking query results using context-aware preferences. In *First International Workshop on Ranking in Databases (DBRank 2007), Istanbul, Turkey*, pages 269–276, Los Alamitos, April 2007. IEEE Computer Society Press.