EXPERTISE MATCHER: INTEGRATING MULTIPLE EXPERTISE INDICATIONS TO RETRIEVE EXPERTS

Ping Liu[a]
Peter Dew[b]

[a,b]School of Computing,
University of Leeds, UK
[a] pliu@comp.leeds.ac.uk
[b] dew@comp.leeds.ac.uk

Session A-4

Abstract

The knowledge and expertise of employees is the major asset that knowledge-based organizations hold. In order to make use of this asset, it is necessary to enhance the communication between employees so that they can share knowledge and expertise. Expertise matching – locating experts with the specific expertise is receiving more and more attention since it facilitates people connection to each other. We develop a conceptual model to integrate multiple expertise indications from heterogeneous data sources and an expertise profile is then created based on this integrated information. This increases the chance of finding experts whilst supports users in selecting the appropriate experts by providing more detailed information of each expert. In addition, an expertise profile is extended to include both keywords form and concepts form based on the domain ontology. This combines the flexibility of keyword search and accuracy of concept search.

Keywords: Knowledge Management, Organizational Memory, Expertise Matching, RDF, ontologies.

Expertise Matcher: Integrating Multiple Expertise Indications to Retrieve Experts

Ping Liu and Peter Dew

School of Computing
University of Leeds, UK
{pliu, dew}@comp.leeds.ac.uk

Abstract

The knowledge and expertise of employees is the major asset that knowledge-based organizations hold. In order to make use of this asset, it is necessary to enhance the communication between employees so that they can share knowledge and expertise. Expertise matching – locating experts with the specific expertise is receiving more and more attention since it facilitates people connection to each other. We develop a conceptual model to integrate multiple expertise indications from heterogeneous data sources and an expertise profile is then created based on this integrated information. This increases the chance of finding experts whilst supports users in selecting the appropriate experts by providing more detailed information of each expert. In addition, an expertise profile is extended to include both keywords form and concepts form based on the domain ontology. This combines the flexibility of keyword search and accuracy of concept search.

Keywords: Knowledge Management; Organizational Memory; Expertise Matching; RDF; ontologies
Suggested track: A Managing organizational knowledge and competence

1 Introduction

Knowledge is the most critical asset for a company (Grant, 1996). Expertise, a major component of tacit knowledge, is the most important basis for the generation of new knowledge, therefore it is the most valuable knowledge (Wilson and Fredericksen, 2000). Expertise defines the organization's unique capabilities and core competencies (Finley, 2001; Holloway, 2000; Olson and Shaffer, 2002). However, if the expertise of employees remains hidden in the individuals' heads and cannot be accessed by others when they need it, then the potential of expertise will be lost (O'Dell et al., 1998). To exploit the great value of expertise it is necessary to enhance the communication between people due to the difficulties in codifying expertise (Stenmark, 2001). Expertise matching - locating experts with the specific expertise, plays an important role in facilitating people connection to each other.

This paper analyses the problem of expertise matching and presents a RDF-based solution to the problem. We employ a conceptual model to integrate multiple expertise indications from heterogeneous data sources and an expertise profile is then created based on this integrated information. This increases the chance of finding experts whilst supports users in selecting the appropriate experts by providing more detailed information of each expert. In addition, an expertise profile is extended to include both keywords form and concepts form based on the domain ontology. This combines the flexibility of keyword search and accuracy of concept search. Our approach has been tested through an experiment, the initial experimental results show the improved performance of expertise matching against traditional database approach.

This paper is organized as follows. It begins with an analysis of the expertise matching problem in Section 2. Section 3 describes the RDF-based expertise matcher and presents the architecture of expertise matching based on integration of heterogeneous data sources. Section 4 briefly presents some experimental results which compare expertise matching performance of our approach and traditional database approach. Finally discussions and conclusions are given in Section 5.

2 Analysis of the problem

Having noticed the importance of finding experts, many organizations have built up systems to facilitate users in locating experts with specific expertise. We have conducted a survey of expertise matching systems among 27 universities. Most of them are UK universities; some of them are US universities. From the survey it can be seen that expertise matching systems are not mature enough due to the following limitations.

- *Manual data collection*: data collection depends heavily on experts or administrative officers.
- *Limited information of experts*: only limited information is stored in a database, which influences the quality of the expertise profile.
- *Unable to rank experts based on their expertise level*: A Boolean search is conducted in most systems and the experts are listed according to alphabetical order rather than their expertise level. Therefore it is difficult to find out who is the most experienced expert.

- *Difficulties in maintaining the up-to-date information*: people's expertise is changing but static database approach cannot catch this change automatically.

Expertise, as one kind of tacit knowledge, has the inherent characteristic of tacit knowledge – it is difficult for people to write down their expertise, they know but unable to express. Although expertise is embodied and embedded, it can often be observed through tangible results (Stenmark, 2002). We use *expertise indications*[1] to refer these tangible results. Different systems use different expertise indications to locate experts. Answer Garden (Ackerman and McDonald, 1996) and ContactFinder system (Krulwich and Burkey, 1996) use *answers to others questions* as expertise indication; Yenta (Foner, 1997) and Know-who email agent (Kanfer et al., 1997) use *emails* as expertise indication; MEMOIR system (Pikrakis et al., 1998) uses *browsing behaviour* as expertise indication; Expert Finder (Mattox et al., 1999), Expertise Finder (Crowder et al, 2002), and Expert Recommendation (Yukawa and Kashara, 2001) use *publications* as expertise indication; CKBS (Liao et al., 1999) uses *projects* as expertise indication; Knowledge pump (Glance et al., 1998) uses *recommendations* as expertise indication, and so on.

There are some slight differences of importance between these expertise indications. However, they are complementary to each other. Hence, an expertise matching system should include as many expertise indications as possible. In most systems only one type of indication is used to create an expertise profile. In order to achieve a more accurate expertise model, there is a need to exploit the multiple expertise indications from heterogeneous data sources. Another limitation of current expertise matching systems is that expertise representation. An expertise profile is either keyword based or concept based. The advantage of keyword-based expertise profile is that experts can be ranked according to their expertise level, however syntax search may bring some irrelevant experts into the results. The advantage of concept-based expertise profile is that the search is more accurate, but it cannot distinguish one expert from another based on expertise level. The third limitation is that output presentation is very weak in most of the existing systems. Normally only personal contact information is provided. Ranking is not very useful for the users if there is no detailed information of each expert, and there is always a degree of "noise" in the ranked list of experts returned by the system. After all it is users who select the

[1] The difference between expertise indicator and expertise indication is that the former refers to terms or phrases reflecting expertise whilst the latter refers to the evidence of expertise such as document authorship.

appropriate experts. The following section describes our approach to solving these problems.

3 RDF-based Expertise Matcher

Expertise matching process can be divided into three main stages. In the *Acquisition* stage, multiple expertise indications from heterogeneous data sources in an organizational memory are captured. In the *Modelling* stage, expertise indicators are extracted from indications and organized in an expertise profile. In the *Retrieval* stage, experts with the required expertise are identified and ranked, and the supported information of each retrieved expert is presented so that users can select the appropriate experts easily. The limitations stated in the Section 2 influence each stage in the expertise matching process. In order to overcome these limitations, it is necessary to (i) integrate multiple expertise indications from heterogeneous data sources; (ii) combine keywords search and concepts search. Semantic web technologies can be employed to achieve the objectives.

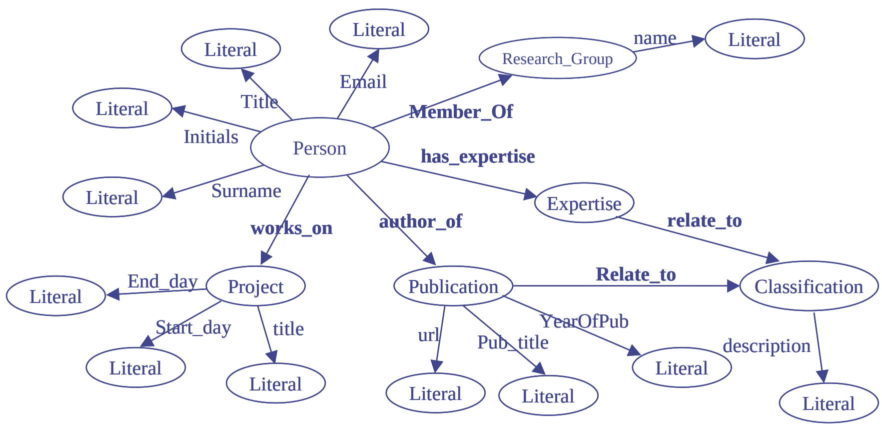

Fig. 1. Sample Conceptual Model used in the RDF-based Expertise Matcher

Note: This is a simplified diagram and hierarchical relationships have not been included due to the space constraints. The major concept in Figure 1 is "Person"; the others are "Publication", "Expertise", "Project", "Research_Group" and "Classification". The relationships between the concepts and the attributes related to each concept are also specified in the conceptual model. For example, a resource of type "Person" may have a property "author_of" whose value is a resource of type "Publication". In the meantime, it can have another property "email" with value "Literal". "author_of" represents the relation between concepts "Person" and "Publication" while "email" represents the attribute related to concept "Person".

We develop a conceptual model (as shown in Figure 1) to integrate heterogeneous data sources. In this conceptual model the shared set of terms in the expertise matching domain as well as the relationships between these terms have been defined. We use XML (Bray et al., 2000) as the data integration language because it provides a uniform syntax (nested tagged elements) to represent heterogeneous data structures. However people are free to add tags and the same information can be represented differently by different XML structures, which brings problems in querying (Berners-Lee, 1998). This limitation is overcome by RDF (Lassila and Swick, 1999) which is based on XML but provides machine understandable semantics for metadata. The RDF data model is just a triple of {subject, predict, object} and therefore it is scalable. Secondly the order of data is not significant and thus querying RDF data becomes easier than querying XML data. The vocabulary of RDF is defined and organized in a typed hierarchy through RDFS (Brickley and Guha, 2000). Furthermore the conceptual model developed can be described by RDFS.

In order to overcome the limitation of keyword search domain ontology is used. Domain ontology characterise the body of knowledge associated with the particular domain such as, the definition of the concepts, the attributes of the concepts (for example, synonyms, abbreviations), and the relations between concepts (for example, is-a and part-of). If experts are searched based on the concepts profiles and ranked based on keywords profiles, then the performance will be improved than using any one type of search alone.

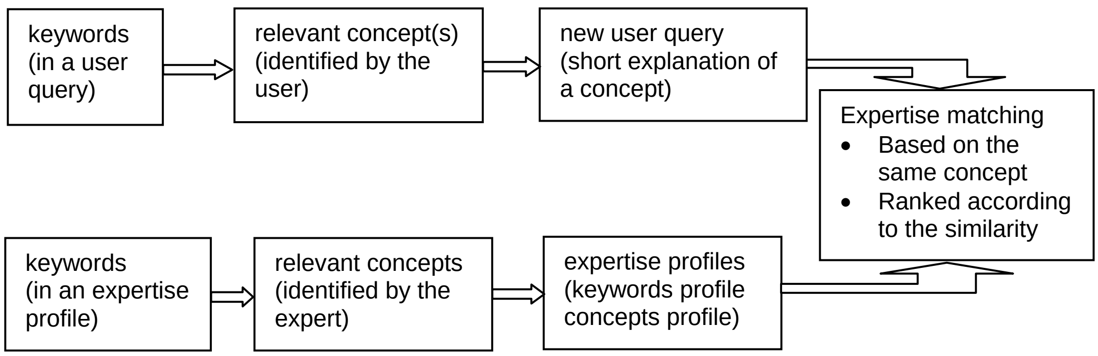

Fig. 2. Matching between user query and expertise profile

Due to the difficulties in automatically linking expertise profile and user queries with concepts, a semi-automatic approach is proposed. As shown in Figure 2, for each concept, a set of keywords is extracted as “relevant keywords”. Based on the expertise profile (a set of keywords with weights), the relevant concepts are retrieved if the description of the concept contains some keywords in the expertise profile. The concept whose description contains the most keywords is listed on the top. Each expert can then confirm if these concept(s) reflect their expertise. Thus, the expertise profile of each expert is built up which includes a set of keywords (with weights) and a set of concepts. A similar process will be applied in confirming the context of user queries.

Once the concepts are selected by the user, the user query is replaced by the short explanation of a concept (in a set of keywords), which will then be used to search for an expert. Only those experts whose expertise profiles include the specified concept are retrieved. The experts are ranked according to the similarity between the keyword profile of experts and the new user query.

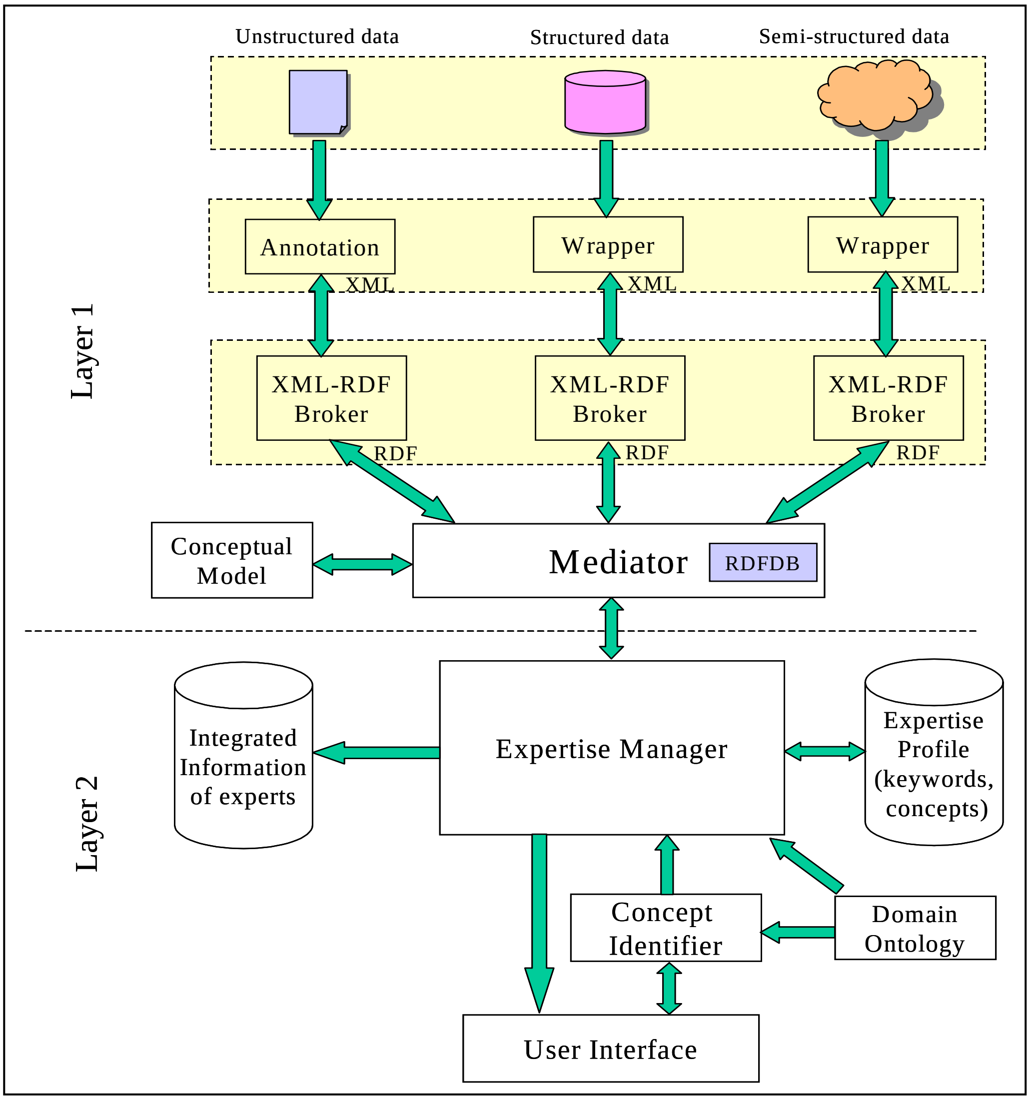

Fig. 3. Architecture of the RDF-based Expertise Matcher

The architecture of the RDF-based Expertise Matcher (as shown in Figure 3) has been proposed which can be divided into two parts, namely, i) semantic information integration; ii) expertise management. The first part was developed based on [Vdovjak and Houben, 2001]. The architecture includes the following components:

- Sources: Contains heterogeneous data sources that include multiple expertise indications.

- Wrapper: Different wrappers such as DB-XML wrappers or HTML-XML wrappers are used to extract relevant information from the original data source and present it to the serialized XML data. For the unstructured data, some manual processes are needed such as adding metadata in XML according to the vocabularies stored in the conceptual model.
- XML-RDF broker: Identifies the relevant concepts in the XML sources and replaces them with the concepts in the conceptual model; the mapping rules are specified in XSLT [Clark, 1999]. This layer also creates the RDF data from the XML instance in order to provide the actual response to a mediator's query.
- Mediator: Maintains the conceptual model. This layer identifies which data sources are relevant to the query, transfers the query to subqueries, and gets subresults from brokers and organizes them according to the conceptual model. The final results (the semantically integrated information of each expert) arrive at the expertise manager.
- Expertise Manager: In addition to maintaining experts' information (experts profiles), expertise manager also creates, stores and retrieves expertise profiles which consist of two forms – keywords and concepts. It receives the extended query and specified concept from the concept identifier and retrieves the experts whose expertise include the required concept. The experts are ranked according to their keyword profiles. The ranked experts with their integrated information are then sent directly to the user interface.
- Concept Identifier: Receives a user's query and provides the relevant concepts according to the domain ontology.
- User Interface: Receives the query from the user and sends the results of the ranked experts together with the detailed information of the experts to the user.

4 Experimental Results

A prototype system has been developed based on this architecture (Liu et al., 2004). This prototype system aims to help PhD applicants locate their potential supervisors. An experiment has been conducted in the School of Computing, University of Leeds to compare the retrieval performance of the RDF-based Expertise Matcher and traditional expertise database approach. The success of the RDF-based Expertise Matcher is measured in terms of whether (i) it saves time in locating experts; (ii) it improves the precision and recall of the search; (iii) it provides richer descriptions of individual experts for selection purposes.

Participants of this experiment were asked to volunteer from the current PhD students in the School. 50% of all the current PhD students attended the experiment. They ranged from 1st year to 3rd year and their research interests were very varied. The same input (their research interests) was given to the two searches and participants were asked to compare the results returned by two searches.

Table 1. Comparison of RDF-based Expertise Matcher and Expertise Database

Fields	**Expertise Database**	**RDF-based Expertise Matcher**
Number of experts retrieved (average)	21.7	4.2
Average time spent on searching (minutes)	8.9	4.6
Precision (average)	22.1%	68.7%
Content information	Limited	Detailed and Satisfied by participants
Ranking	55% not useful; 40% partially useful	100% useful
Recall	Lower	Higher

The detailed experiment design and results can be found in (Liu et al., 2004). Table 1 summarises the comparison of the RDF-based Expertise Matcher and expertise database in six fields[2]. It shows that the RDF-based Expertise Matcher provides better performance than traditional database approach.

5 Discussion and Conclusions

The major strengths of the RDF-based Expertise Matcher are (1) it is based on the multiple expertise indications from heterogeneous information stored in an organizational memory rather than a single data source such as publications; (2) the hybrid approach combines the advantage of flexibility of keyword search and accuracy of concept search; (3) it supports not only “expertise identification” but also “expertise selection” (McDonald and Ackerman, 1998) by providing integrated and detailed information of each returned expert.

[2] Although the number of relevant experts retrieved is known, it is difficult to find all the relevant experts. Therefore we do not give the actual value of recall. However, the total number of relevant experts are the same for both searches, so what is important is which search provides a larger number of relevant experts. The experiment results show that RDF-based Expertise Matcher retrieves more relevant experts (in average) than traditional expertise database. Therefore recall is higher in the RDF-based Expertise Matcher.

Most current knowledge management programmes only focus on gathering, organising, and retrieving information. This work contributes to the other kind of knowledge management - expertise sharing by retrieving experts with the required expertise. This paper discusses how to apply semantic web technology – RDF/RDFS, XSLT, ontologies – to integrate multiple expertise indications from heterogeneous data sources. The evaluation of the prototype system indicates the improved performance of expertise matching traditional database approach. Although this research focuses on academic environment, our approach is applicable to most knowledge-based organizations. We are planning to test our system with more users in various disciplines and get more feedback.

Reference:

Ackerman, M.S. and McDonald, D.W. (1996) Answer Garden 2: Merging Organizational Memory with Collaborative Help. In Proceedings of the ACM Conference on Computer-Supported Cooperative Work (CSCW'96), Boston, MA, 1996, pages. 97 – 105.

Berners-Lee, T., (1998) Why RDF model is different form the XML model (1998) http://www.w3.org/DesignIssues/RDF-XML.html [February 26, 2004]

Bray, T., Paoli, J., Sperberg-McQueen, C.M. and Maler, E. (2000) Extensible Markup Language (XML) 1.0 (Second Edition), W3C Recommendation, October 2000. http://www.w3.org/TR/2000/REC-xml-20001006 [February 26, 2004]

Brickley, D. and Guha, R.V. (2000) Resource Description Framework (RDF) Schema Specification 1.0,World Wide Web Consortium March 2000. http://www.w3.org/TR/2000/CR-rdf-schema-20000327/ [February 26, 2004]

Clark, J. (1999) "XSL Tranformations (XSLT)," W3C Recommendation November, 1999, http://www.w3.org/TR/xslt [February 26, 2004]

Crowder, R., Hughes, G. and Hall, W. (2002) An Agent Based Approach to Finding Expertise. In Fourth International Conference on Practical Aspects of Knowledge Management, 2-3 December, 2002, Vienna, Austria

Finley, M. (2001) FUTURE SHOES: "Expertise Management 101", February 16, 2001. http://www.computeruser.com/articles/daily/7,5,1,0216,01.html [February 26, 2004]

Foner, L.N. (1997) Yenta: A Multi-Agent Referral-Based Matchmaking System. Proceedings of the First International Conference on Autonomous Agents (Agent'97), Marina del Rey, CA, February 1997, pages 301-307

Glance, N., Arregui, D. and Dardenne, M. (1998) Knowledge Pump: Supporting the Flow and Use of Knowledge. In Borghoff, U.M.; Pareschi(eds.):Information Technology for Knowledge Management. Springer Verlag, pages 33-53

Grant, R.M. (1996) "Toward a Knowledge-Based Theory of the Firm," in Strategic Management Journal 17, pages 109-122

Holloway, P. (2000) "How to Capture and Deploy Tacit Knowledge in Your Organization," in Braintrust January 2000. Scottsdale. http://www.knowledgeharvesting.org/kho/presentations.htm [February 26, 2004]

Kanfer, A., Sweet, J. and Schlosser, A. E. (1997) Humanizing the net: Social navigation with a "know-who" email agent. In Proceedings of the 3rd Conference on Human Factors and the Web. Denver, Colorado, June 12, 1997

Krulwich, B. and Burkey, C. (1996) The ContactFinder: Answering bulletin board questions with referrals. In Proceedings of the 1996 National Conference on Artificial Intelligence (AAAI-96), Portland, OR, vol.1, 1996, pages.10-15.

Lassila, O. and Swick, R.R. (1999) "Resource Description Framework (RDF): Model and Syntax Specification". W3C Recommendation, February 1999, http://www.w3.org/TR/1999/REC-rdf-syntax-19990222/ [February 26, 2004]

Liao, M., Hinkelmann, K., Abecker, A., and Sintek, M. (1999) A Competence Knowledge Base System as Part of the Organizational Memory. Proceedings of 5th Biannual German Conference on Knowledge-Based Systems – Survey and Future Directions, March, 1999 in Würzburg, Germany, pages 125-137.

Liu, P., Curson, J. and Dew, P. (2004) Use of RDF for Expertise Matching within Academia, to be published in the Journal of Knowledge and Information Systems

Mattox, D., Maybury, M. and Morey, D. (1999) Enterprise expert and knowledge discovery. In Proceedings of the 8th International Conference on Human Computer Interaction (HCI International'99), Munich, Germany, August. 23-27, 1999; pages.303-307

McDonald, D. W., and Ackerman, M. S.(1998). "Just Talk to Me: A Field Study of Expertise Location," in Proceedings of the ACM Conference on Computer-Supported Cooperative Work (CSCW '98), pages 315-324

O'Dell, C., Jackson, C., Grayson, J. (1998) If Only We Knew What We Know: The Transfer of Internal Knowledge and Best Practice, Free Press1998.

Olson, L. and Shaffer, R. (2002) "Expertise Management – and Beyond". White paper in RGS Associates. http://www.rgsinc.com/publications/pdf/white_papers/Expertise_Management_and_Beyond.pdf [February 26, 2004]

Pikrakis, A., Bitsikas, T., Sfakianakis, S., Hatzopoulos, M., DeRoure, D., Hall, W., Reich, S., Hill, G. and Stairmand, M. (1998) MEMOIR - Software Agents for Finding Similar Users by Trails. In Proceedings of the 3rd International Conference on the Practical Applications of Agents and Multi-Agent Systems (PAAM-98), London, UK, 1998, pages 453-466

Stenmark, D. (2001). "Leveraging Tacit Organisational Knowledge," in Journal of Management Information Systems, Special Winter Issue, Vol. 17, No. 3, pages. 9-24.

Stenmark, D. (2002). Information vs. Knowledge: The Role of intranets in Knowledge Management. In Proceedings of HICSS-35, Hawaii, January 7-10, 2002 Available: http://www.viktoria.se/results/result_files/183.pdf [February 26, 2004]

Vdovjak, R., and Houben, G. (2001) "RDF Based Architecture for Semantic Integration of Heterogeneous Information Sources," in Workshop on Information Integration on the Web (2001) pages.51-57

Wilson, L.T. and Fredericksen, D. (2000) "Harvesting eProcess Know-how," in The Delphi eBusiness Summit 2000, 9 May 2000, Coronado, California. http://www.knowledgeharvesting.org/kho/presentations.htm [February 26, 2004]

Yukawa, T. and Kashara, K. (2001) An Expert Recommendation System using Concept-based Relevance Discernment. 13th IEEE International Conference on Tools with Artificial Intelligence (ICTAI'01) November 7-9, 2001, Dallas, Texas, pages. 257-264