# Exploring RDF for Expertise Matching within an Organizational Memory

Ping Liu, Jayne Curson, and Peter Dew

Informatics Research Institute, School of Computing, University of Leeds
Leeds, LS2 9JT, United Kingdom
{pliu,dew}@comp.leeds.ac.uk
j.m.curson@leeds.ac.uk

**Abstract.** Organizations have realized that effective development and management of their organizational knowledge base is very important for their survival in todays competitive business environment. People, as a special knowledge asset, also attract the interest of many researchers because, only through people communicating with one another, can they really share their tacit knowledge and skills that can be more valuable than explicit documentation. The need to be able to quickly locate experts among the heterogeneous data sources stored in the organizational memory has been recognized by many researchers. This paper examines the advantages of using RDF (Resource Description Framework) for Expertise Matching. The major challenge is to semantically integrate heterogeneous data sources stored in the organizational memory and facilitate users to locate the right people. We present a practical application of this using a case study where PhD applicants can locate potential supervisors before they formally apply to a university.

## 1 Introduction

In the current economic environment, organizations have realized that effective development and management of an enterprise's organizational knowledge base will be a crucial success factor in the knowledge-intensive markets of the current century [1]. An Organizational Memory[1] is designed to store what employees have learned from the past in order for it to be reused by current employees in solving problems more effectively and efficiently. There are two kinds of retrieval in the organizational memory. One is "*information retrieval*" which aims to provide the knowledge required by the task at hand. However, access to information only is not sufficient, people often need to communicate with each other in order to find more important information which cannot be obtained from explicit documentation. That is

[1] In this paper, Organizational Memory is synonymous to Organizational Memory Information System because we focus on the technical aspects of OM. See [1] for a definition of OMIS.

A. Banks Pidduck et al. (Eds.): CAISE 2002, LNCS 2348, pp. 100-116, 2002.

why we need another kind of retrieval – "*people retrieval*". The process of finding relevant people who have similar interests is also called Expertise Matching. As noted by many researchers [2,4,6,9,15,25,36,39,40], employees learn more effectively by interacting with other employees because the tacit knowledge and expertise people possess are difficult to codify and store in a knowledge management system. There is widespread agreement that the highest-value knowledge is the tacit knowledge stored in peoples heads [27]. Consequently, we put our research emphasis on how to support "people retrieval" rather than "information retrieval".

If users wish to search information in web pages they can use search engines. However, if they want to locate somebody with the required expertise, there is no existing system which provides a satisfactory result. Users have to manually check different data sources stored in the organizational memory in order to find pieces of information relevant to an expert and then combine them manually. Considering the huge amount of information that the organizational memory stores, it is no surprise that searching for people with specific expertise is a common problem in nearly every company [31]. The main challenge addressed in our work is that of how "people retrieval" can be improved by extracting relevant information associated with an expert from different data sources and semantically integrating them.

This paper is organized as follows: It begins with an analysis of the Expertise Matching problem in Section 2. Section 3 describes the possible approaches and justifies the use of RDF in our solution of the Expertise Matching problem. Section 4 demonstrates our solution of Expertise Matching in a Brokering System which is currently being developed at the University of Leeds to help PhD applicants locate potential supervisor(s). It also describes the rationale for the system and presents the architecture. The use of the system is illustrated in Section 5 along with the key results. Finally, Section 6 compares our work with other related research and indicates areas that require further investigation.

## 2 Analysis of the Problem

There are many definitions of expertise. One definition from Webster's dictionary is "processing special skill or knowledge; trained by practice; skillful or skilled" [23]. Bedard gave a similar definition, "a combination of knowledge and ability, and the capability to achieve results with this knowledge" [5]. It is also defined as "a process by which individuals develop the ability to achieve task-specific superior performance" [32] and "the ability, acquired by practice, to perform qualitatively well in a particular domain" [22]. However, the substance of skills, knowledge and ability is a hidden variable and difficult to codify. This is why databases such as COS[2], VTED[3], BATH[4], New England[5] express expertise in terms of several keywords. REPIS[6] is distinct from these and a brief description of the system is given here. The

[2] COS Expertise http://expertise.cos.com/dics/expfields.shtml
[3] The Virginia Tech Expertise Database http://www.rgs.vt.edu/vted/
[4] University of BATH Directory of Expertise http://www.bath.ac.uk/expertise
[5] University of New England Expertise Search http://research.une.edu.au/
[6] University of Leeds Research Expertise and Publication Information System http://repis.leeds.ac.uk

University of Leeds Research Expertise and Publications Information System (REPIS) is a web-based information management system. It stores information about publications and research projects acquired from a variety of different sources. The principal objectives of REPIS are to provide a directory of research expertise across the University and to provide an introduction to the University's research activities for potential collaborators in academia, industry, government and charities. The REPIS Expertise Matcher acts as a *knowledge broker* connecting knowledge seekers and knowledge providers as shown in Figure 1. The difference between REPIS and other systems is that expertise is not input by the individual academics themselves but derived from their associated work outputs, in other words, their publications and projects. The current REPIS system uses search methods employed by SQL Server 2000 to search publication and project databases in order to locate the most appropriate expert(s).

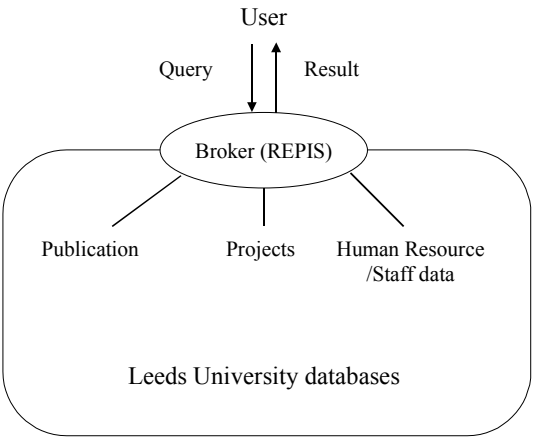


**Fig. 1.** Expertise matcher as a knowledge broker linking users and experts

There are limitations associated with DBMS techniques. Firstly, users looking for experts in a particular field need a lot of information in order to assess if this is the right person to contact. For example, users need to know the experts' position(s), their research interests, the project(s) they are working on or have worked on in the past, records of activities, and they may even want to read research papers they have produced. Manually creating a database to store all this information is very difficult and expensive. Secondly, there is the critical problem of maintaining up-to-date information. A person's expertise changes over time and it is not feasible to rely on the individual to report developments to their expertise profile and even so, the database maintenance task would be significant if many individuals were involved, for example, there are nearly 4000 academic related staff at the University of Leeds.

One important question is: "The information required is stored in the organizational memory, is it possible to automatically extract the relevant information from disparate data sources and integrate them?". To answer this question, it is necessary to examine closely what type of information is often stored in the

organizational memory. There are a number of different data sources varying from structured data (such as databases), semi-structured data (such as web pages), to unstructured data (such as text files). This heterogeneity brings many difficulties in knowledge sharing. Busse [12], Seligman [34] and Sheth [35] present different classifications of heterogeneity, which can be summarized into 5 types, (1) *Heterogeneous interfaces;* (2) *Heterogeneous attribute representations;* (3) *Heterogeneous schemas;* (4) *Heterogeneous semantics;* (5) *Object identification*.

## 3 Approaches to Solving the Heterogeneous Problems

The problems of heterogeneity have been addressed in various projects and corresponding techniques have been proposed. Traditional approaches, which include standards like ODBC, middleware, federated database system (FDBS), and mediator-based information system, all suffer certain limitations. For example, middleware may be *costly* and may be *inefficient* compared to using native interfaces [34]. FDBS is only *applicable* to databases whilst mediator-based information systems require the software developers to have a clear understanding of a variety of metadata, as well as a comprehensive understanding of schematic heterogeneity [35]. In rule-based mediators, rules are mainly designed in order to reconcile structural heterogeneity [24], whilst for the reconciliation of the semantic heterogeneity problems, the semantic level also has to be considered [37]. The literature on integration is more concentrated on syntax and structure with few people focusing on semantic interoperability (see for example [21], [37]).

Extensible Markup Language (XML) [10] is accepted as the emerging standard for data interchange on the web. XML has defined a neutral syntax that can transform diverse data structures into graph-structured data as nested tagged elements [34]. In this way, heterogeneous data structures can be represented in a uniform syntax – XML. Using XML, three problems listed above can be alleviated, i.e. heterogeneous DBMSs, heterogeneous attribute representations, and heterogeneous schemas. However, XML cannot support integration at the semantic level. For example, there are two expressions: <Surname> Black </Surname> and <Lastname> Black </Lastname>, which seem to carry some semantics. However, the system does not understand that Surname and Lastname mean the same thing and that they are related to another concept, for example, “Person”. XML Schema provides support for explicit structural cardinality and data typing constraints, but does not provide much support for the semantic knowledge necessary to integrate information [28]. Again, XML does not play a very significant role in object identification.

RDF (Resource Description Framework) [30] and RDFS (the Schema Language for RDF) [11] are W3C recommendations for describing metadata on the web. They can be used to solve the semantic heterogeneous problem. RDF provides a standard representation language for web metadata based on directed labelled graphs [29]. It consists of three object types: Resource, Property and Statement. Every resource has a Uniform Resource Identifier (URI). The use of URIs to unambiguously denote objects, and of properties to describe relationships between objects, distinguish it fundamentally from XML’s tree-based data model [19]. The same RDF tree can be

expressed differently in many XML trees because the order of elements in an XML document is very meaningful. So RDF successfully avoids the problem of querying XML trees which attempts to convert the set of all possible representations of a fact into one statement [7].

In order to solve the heterogeneous semantics problem, a shared set of terms describing the application domain with a common understanding is needed. Such a set of terms is called an ontology or a conceptual model[7], which includes not only the definition of the terms, but also the relationships between these terms. The most important role for RDFS is to define the ontology [28]. RDFS enables the interpretation of RDF descriptions. Through using ontologies to make the implicit meaning of their different terminologies explicit, it is then possible to dynamically locate relevant data sources based on their content and to integrate them as the need arises [16].

Having justified the importance of RDF/RDFS for semantic information integration, the use of RDF/RDFS in an organizational memory is now being explored. The aim is to provide a coherent and meaningful view of the integrated heterogeneous information sources associated with each particular expert. In the next section, this is illustrated through a practical application, namely a brokering system, which matches PhD applicants with potential supervisors in the School of Computing at the University of Leeds.

## 4 Experiment and Rationale

The School of Computing in the University of Leeds is a large department which each year attracts approximately 50 applications from potential research students. Potential research students can either trawl through web pages and search databases to try and locate information about potential suitable supervisors or they may simply ask the School's PhD Admissions Tutor to select a suitable supervisor for them based on their proposed research topic. The problems are (1) It is still very difficult for the PhD Admissions Tutor to recall up-to-date details of all the expertise and research interests for each academic as individual expertise and research interests may continually change and develop. (2) The PhD Admissions Tutor may not fully understand the applicants' intents because some applicants use quite specific technical terminology. As a result, the supervisor that the PhD Admissions Tutor recommends may not be the most suitable, and there exists a real possibility that some appropriate applicants are rejected because their needs cannot be appropriately matched in this way.

The design of our Brokering System aims to improve the process of matching supervisors and potential research students by enabling the potential applicant to make more informed choices about their supervisor before they formally apply to the University and benefiting both the School and the applicant in the long-run.

---

[7] The difference between Ontology and Conceptual Model is that "Ontology is external to information systems and is a specification of possible worlds in some particular domain that covers multiple and often a priori unknown information systems while a conceptual model is internal to information systems and is a specification of one possible world of that domain" [8].

### 4.1 User Study

To identify the support tasks needed in this Brokering System, let us consider the following scenario, which represents a typical case of the problem described above:

> *Mary is a Masters student at the University of Manchester and plans to study for a PhD. She searches the web pages of several universities, including the University of Leeds; her preferred research interest is "heterogeneous database systems". Mary first navigates the School of Computing website at the University of Leeds and browses the homepage of each member of staff. She quickly finds that there are a large number of staff in the School and many of whom are not active researchers. Then she decides to browse the research groups in order to quickly locate a potential supervisor. She finds these websites are not well organized. Although she searches very carefully, she still does not find an academic who can match her requirements. She thinks that maybe there are no academics conducting research in this area and she should give up applying to Leeds University.*

This is not the desired outcome as there are people who could supervise her at Leeds. The scenario draws attention to the following problems involved in identifying the potential supervisor(s):

- *Low recall*: This means that some relevant people are missed. This is mainly due to: (1) There is a large number of staff in the School and it is a very time consuming task for the user to access each person's homepage; (2) The web page of each research group does not give detailed information on the individuals in the group. As a consequence, the user may not find the relevant person even when searching carefully.
- *Low precision*: This means that some of the people found are not experts in the preferred research area. It is not always the case that researchers working in the same research group have very similar research interests or expertise. Users still need to conduct further assessment by looking carefully at the detail of each researcher in order to determine if that individual is a suitable supervisor. Therefore, the number of real experts is very small compared to the total number of people retrieved.

The following is the ideal situation that Mary wants the system to provide:

> *When Mary conducts a search by entering her research interests, several relevant research areas are returned. Mary chooses "Information Integration and Databases" as her preferred research area, and two researchers are displayed. Each researcher has his/her own detailed information including research interests, the projects they are working on or have worked on, the papers they have published, and the technical reports which can be downloaded. Mary compares these two researchers and reads abstracts of 2 papers, she then chooses one of the two to be her preferred supervisor and starts completing the application form.*

From the ideal situation above we can identify the most significant support tasks required of the Brokering System:

- Understanding user needs/identification of expertise requirements;
- Understanding the domain knowledge in order to provide translation between researchers' expertise and user interests;
- Providing an integrated view to the user from the different/diverse information sources;
- Capturing changes to the expertise profile of researchers.

### 4.2 The Proposed Architecture of Our Brokering System

The architecture for a typical Brokering System is described in [38]. This architecture has been adopted here and can be divided into five separate layers as shown in Figure 2 below. Figure 2 also illustrates the different data sources used in our case study.

- *Source Layer*: Contains data sources that are relevant to identifying the expertise of each potential supervisor such as personal homepages which includes personal contact information, research interests, associated research group(s), and recent publications; the REPIS database which stores information about publications and projects across the University; and technical reports which are online documents stored in the School of Computing database. These data sources are built by different people for different objectives or different users, some of the data across these three data sources is duplicated. For example, information on a particular publication authored by a member of staff may be stored in all these data sources.
- *XML Instance Layer*: Presents the serialized XML data transferred from the original data sources. This is through DB-XML wrappers or HTML-XML wrappers. For these unstructured data, some manual processes are needed such as adding metadata in XML according to the vocabularies stored in the Conceptual Model.
- *XML2RDF Layer*: 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 [13]. These mapping rules are defined by the application designer and can be modified if the concepts of the source change. However, the underlying Conceptual Model should be stable as it is the basis for the semantic integration; if it has to be changed, then the RDF model and the mapping rules should be modified accordingly. This layer also creates the RDF data from the XML instance in order to provide the actual response to a mediator's query.
- *Mediator Layer*: Maintains the Conceptual Model (shown in Figure 3). This layer identifies which data sources are relevant to the query, transfers the query to subqueries, and gets subresults from brokers. These subresults are input into RDFDB[8], and through searching RDFDB, the final results arrive at the application layer.
- *Application Layer*: Receives the query from the user and produces a result to the user.

[8] An RDF database http://web1.guha.com/rdfdb/

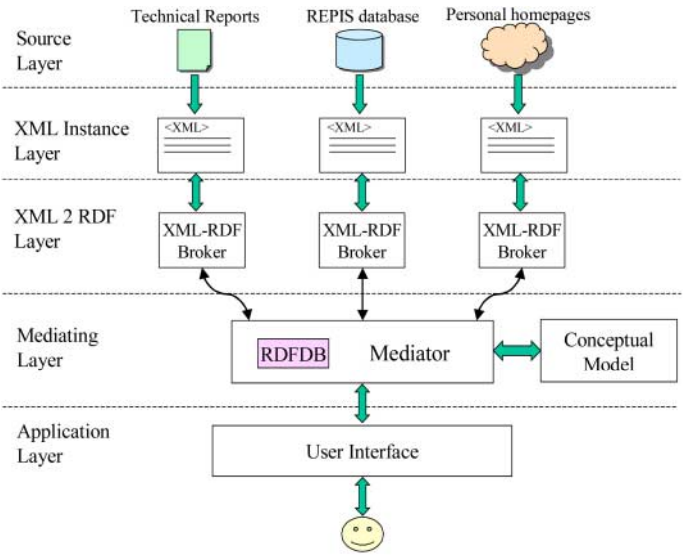


**Fig. 2**. Architecture of integration of heterogeneous information sources

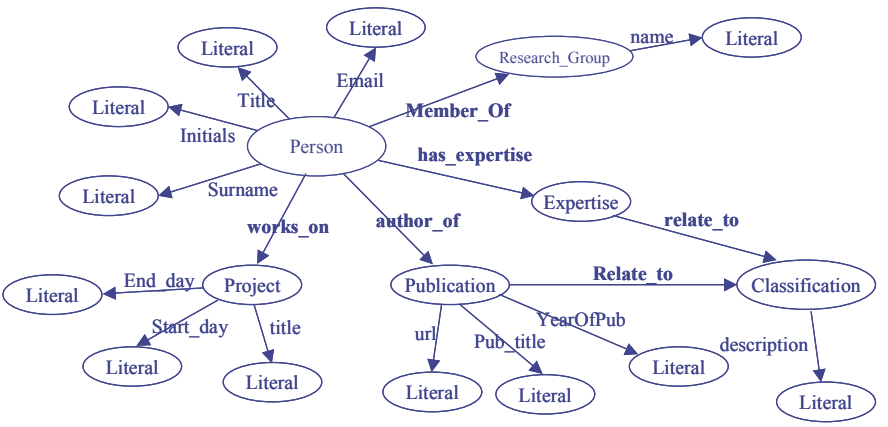


**Fig. 3.** Sample conceptual model used in the brokering system

Note: This is a simplified diagram and hierarchical relationships have not been included due to the space constraints. An example of the underlying hierarchical structure associated to concept "Person" is given in Figure 4. The major concept in Figure 3 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".

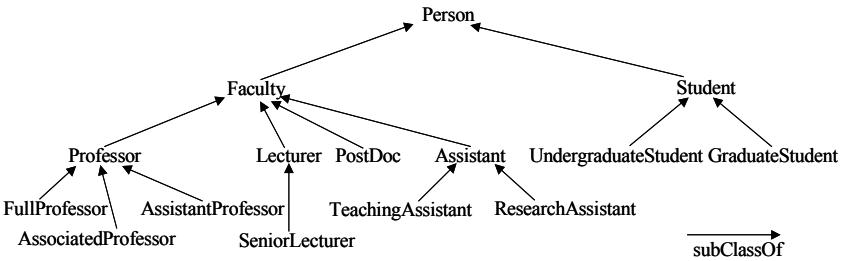


**Fig. 4.** Example of underlining hierarchical structure associated with the concept "Person"

### 4.3 Brief Implementation Details

The implementation of the architecture described in the previous section includes several crucial aspects as follows:

1. *Indexing and retrieval of concepts:* The actual concepts and their associated keywords and supervisors are stored in a relational database. This database is connected to the Java system code via JDBC. The possible relevant concepts are retrieved based upon the research interests that the user inputs.
2. *Constructing the detailed information for supervisors:* Firstly, relevant information from diverse data sources should be collected. The information stored in the web pages and the REPIS database is transferred into XML form using wrappers. Some manual annotations are needed for interpreting the information stored in the unstructured data sources. Secondly, these XML files are then transferred into RDF data according to the mapping rules specified in XSLT. Thirdly, the separate RDF data is input into an RDF database -- RDFDB. Fourthly, a search is conducted on RDFDB to produce the complete detailed information for each supervisor. Duplicate information is removed at this step. The third and fourth steps are implemented through a Java interface for RDFDB.
3. *Ranking the expertise of potential supervisors:* Individual supervisor's expertise profile is represented as vectors of keywords based upon the searching results from RDFDB. Each keyword is assigned a weight using tfidf metric [3]. The relevance of each potential supervisor is calculated through the similarity between the profile of each potential supervisor and the set of keywords used to describe the main concepts. The weight attributed to each potential supervisor is then converted into a percentage value by dividing the weight attributed to the individual by the sum of the weights of all the potential supervisors.
4. *Displaying the semantically integrated information of potential supervisors:* This is also implemented using Java programming. The search results from RDFDB are firstly constructed to a XML file, and then into an HTML file which is presented to users through XSLT.

### 4.4 Data Validation

In order to provide high quality information about each expert, it is unavoidable to rely on accurate information being available. For example, personal homepages and technical reports should be updated annually. In our case, the REPIS database, as a core data source, is heavily relied upon. This is because the data stored in REPIS on individual academics has been validated by the administrator of each department. There are also a number of automated validation processes built into REPIS. For example, one data source held in REPIS is ULRICHs[9] the authorative serials bibliographic database providing details of title and the International Standard Serial Number (ISSN) for journals published throughout the world. If an administrator tries to input details for a publication type of 'academic journal paper' and indicates an incorrect journal and/or ISSN then they will be automatically informed of this and provided with the correct title and/or ISSN details. The other data sources (such as personal homepage and technical reports) are complementary to REPIS in order to provide a richer description of each expert.

## 5 System Walk through and Key Results

A prototype brokering system, using the architecture described above, has been built and used to match PhD applicants with potential supervisors. The search for potential supervisor(s) follows 3 steps which are described below:

1. The user inputs a description of preferred research interest(s) and selects individual research areas which are the most relevant.
2. The user views the name of each academic working in the relevant research area.
3. The user views the detail of the preferred supervisor.

**Step1**: Initially the user inputs a brief description of their general research interests. This description is formulated in natural language. A list of relevant research areas will then be displayed (Figure 5). The selection of the relevant research areas is on the domain ontology which is a combination of ACM Computing Classification[10] and the computing dictionary[11]. The relevant research areas are ranked according to the number of keywords contained in the resarch interest field entered by the user which are relevant to each research area. Each result consists of three parts. First, the value which indicates the number of keywords that the user inputs which are relevant to the research area; second, the research area which is displayed in upper case; third, a list of the relevant keyword stems which are used to search all variants of the same keyword. The user can view the detailed information of each research area by clicking on "Show me the detail" or they can "Accept" the research area if they feel this is an area in which they would like to conduct research. They may accept as many research areas as they wish.

[9] ULRICHs http://www.ulrichsweb.com/
[10] ACM Computing Classification http://www.acm.org/class/1998
[11] Online computing dictionary http://foldoc.doc.ic.ac.uk/foldoc/index.html

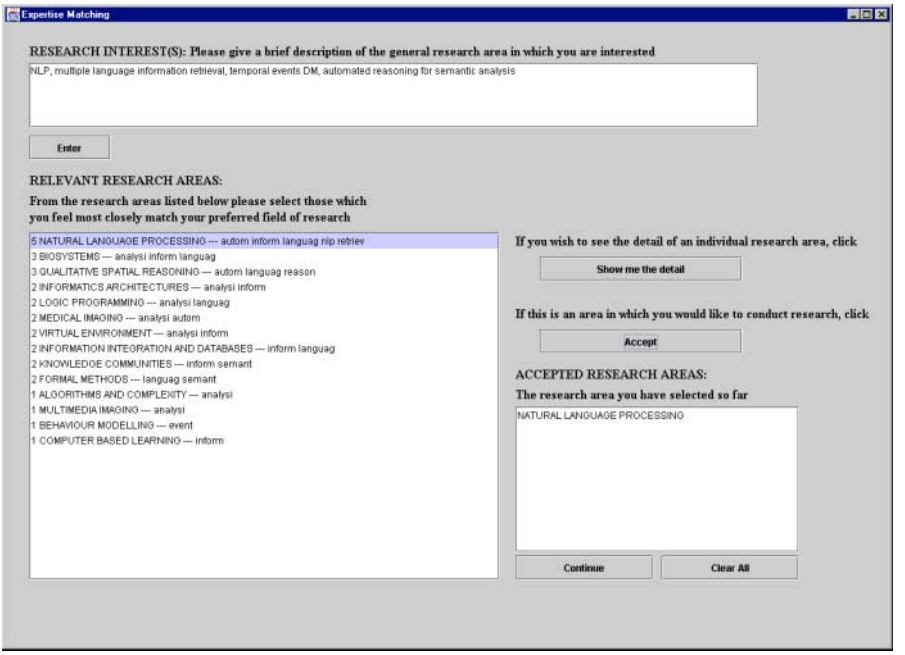


**Fig. 5.** Step 1 user interface for inputting research interests


**Fig. 6.** Step 2 display the potential supervisor(s) for each preferred research area selected

**Step2**: The user can select any relevant reseach area in order to view a list of potential supervisors working in that research area (as shown in Figure 6). The potential supervisors are ranked according to how likely it is that this person will be selected as the potential supervisor. The example shown in Figure 6 indicates that there is a 64%

chance of choosing Mr E Atwell as the potential supervisor, a 23% chance for Dr. D. C. Souter and a 13% chance for Dr. L. W. Bod. The technique used to calculate the possibility for each potential supervisor being chosen is based on the Vector Model of information retrieval [3]. The detailed process of calculating this is outside the scope of this paper; a similar algorithm has been developed by [18].

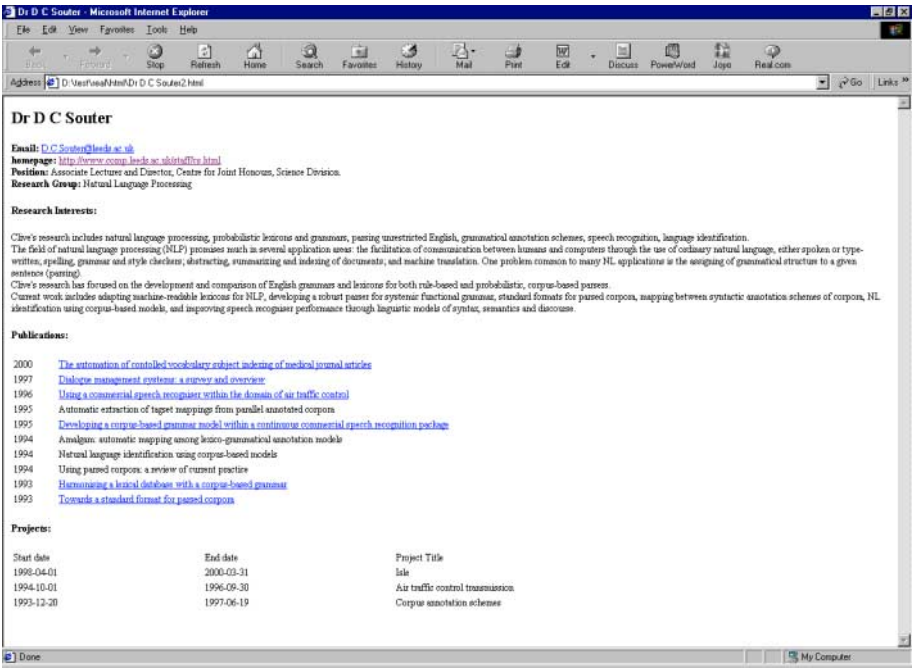


**Fig. 7.** Step 3 detailed information on the selected potential supervisor

**Step 3:** The complete personal profile of the particular potential supervisor (as shown in Figure 7) will be displayed if the user clicks on "View supervisor". The full detail page of Dr. D. C. Souter appears like a standard personal homepage currently existing in the School of Computing, but when we take a closer look at it, we find that it includes information taken from different data sources. As shown in Figure 7, the data is retrieved as follows: (1) The personal contact information and research interests are retrieved from the personal homepage; (2) The publication section is a combination of information from the personal homepage, from a series of technical reports which can be downloaded from the REPIS database. The duplicate information is deleted and the final results are reorganized so that the user is not aware where this information comes from; (3) The project information is also retrieved from the REPIS database.

The prototype system has been tested using the application forms of 28 current PhD students. For each applicant, the research interests from each application form were input into the research interests field of our Brokering System. The relevant research areas were then selected and the potential supervisors selected by through checking the detail of the individuals who were working in these areas. The initial results were then checked to determine whether the names of their actual supervisors

were in the final lists of accepted potential supervisors. In order to find out if there is a benefit when the potential supervisors are ranked according to how likely it is that each potential supervisor will be selected, the prototype system was tested again using the same application forms after the ranking function had been added. The initial testing results are shown in Table 1; in particular, we can highlight the following:

**Table 1** Initial results of searching for a potential supervisor

| Number of application forms | Number of potential supervisors | How far down the list is the actual supervisor placed (before ranking)? | How far down the list is the actual supervisor placed (after ranking)? |
|---|---|---|---|
| 4 | 1 | 1 | 1 |
| 3 | 2 | 1 | 1 |
| 2 | 2 | Join supervisor both 1 and 2 | Join supervisor both 1 and 2 |
| 5 | 2 | 2 | 1 |
| 1 | 3 | 2 | 1 |
| 5 | 3 | 3 | 1 |
| 2 | 4 | 3 | 1 |
| 1 | 4 | 4 | 1 |
| 2 | 5 | 3 | 3 |
| 1 | 5 | 1 | 1 |
| 1 | 6 | 1 | 1 |
| 1 | >7 | variable | variable |

The names of the actual supervisors of 27 of the PhD students were shown in the final lists of the accepted potential supervisors. Only in one case was it difficult to ascertain if the actual supervisor was in the final list. This is because the applicant indicated many research interests on his application form, and as a result, lots of researchers could potentially serve as his potential supervisor which makes the selection more difficult.

It is noted that after the ranking function has been added, the names of the 14 PhD students' supervisors were listed higher than before. Although we cannot say for certain that the actual supervisor of each student is the most appropriate supervisor, it should be noted that the supervisor of each student is selected manually and methodically by the students themselves or by the PhD Admissions Tutor. This means that if the names of the actual supervisors are placed at the top of the results list most of time then the system is considered to be successful.

## 6 Conclusion and Directions of Future Work

The strengths and weaknesses of the work reported here is compared with several related projects concerned with searching web documents, searching XML-based documents in an organizational memory, and searching for people in organizations.

1. Searching web documents: SHOE (Simple HTML Ontology Extensions language) project [26] is one of the earliest trials on adding semantic information to web pages according to predefined vocabularies or ontology. On2broker [21] also provides languages to allow web page authors to annotate their web documents with ontological information. It allows users to access information and knowledge from the web and infer new knowledge with an inference engine.
2. Searching XML-based documents in an organizational memory: Osirix [33] tries to solve the heterogeneous data source problem in organizational memory using XML technology. Annotated XML documents are created and then searches are conducted on these XML documents.
3. Searching for people in organizations: CKBS [31] system, as part of organizational memory, builds upon an ontology-based model of competence fields. Expert Locator [14] uses concepts, a large pre-built, technical thesaurus as the initial ontology and enhances it using simple AI techniques. This results in a reduction of the ambiguity of the single keyword problem and also exploits domain knowledge for the search. Knowledge maps [20] make expertise accessible through visual interface based on a common framework or context to which the employees of a company can relate.

The major challenge of our system is that the different technologies from different research areas have been integrated and applied to the Expertise Matching problem. Firstly, our Brokering System does not restrict the data source to webpages in the case of SHOE and On2broker. Secondly, it also provides the semantic interoperability which is not addressed in Osirix system. Thirdly, it makes full use of all the information stored in the organizational memory and provides dynamically updated information about each person, which is richer than CKBS and Expert Locator. Fourthly, it can be used by both internal and external users rather than the employees of a company as in the case of Knowledge maps. The major advantages of using RDF for Expertise Matching are its abilities to: (1) integrate the pieces of information from the organizational memory to form a new up-to-date profile of each expert; and (2) improve the quality of the information through removal of duplicate data.

In this paper, we explore the use of RDF in matching PhD students and their potential supervisors. The same technology will be used in the KiMERA[12] project which aims to help people from industry to locate relevant experts in academia and facilitate collaboration. The internal users, who are employees of the University of Leeds, can also benefit from finding people who are doing similar things and exchange tacit knowledge. To date our prototype system has been tested using 28 PhD application forms. The initial results are very promising. Further user experiments will be conducted in the near future in order to evaluate our system against two main criteria: (1) How accurate the results are in terms of finding the right experts; (2) Whether there is any benefit to using semantic web technology (such as RDF) in terms of improving Expertise Matching performance.

[12] The Knowledge Management for Enterprise and Reachout Activity http://kimera.leeds.ac.uk

## Acknowledgements

We wish to thank Dr Richard Drew from Symularity Ltd for helpful advice on the Brokering System and for his work contributing to the REPIS database. We also thank Dr Vania Dimitrova for helpful comments on the text.

## References

1. Abecker, A. and Decker, S. Organizational Memory: Knowledge Acquisition, Integration, and Retrieval Issues in Knowledge-Based Systems, Lecture Notes in Artificial Intelligence, Vol. 1570, Springer-Verlag, Verlin, Heidelberg, pages 113-124, (1999)
2. Ackerman, M. S. and Halverson, C., Considering an Organization's Memory, in Conference on CSCW'98 pages 39-48, ACM Press, Seattle, WA, (1998)
3. Baeze-Yates, R. and Ribeiro-Neto, B., Modern information retrieval Imprint Addison-Wesley Longman (1999)
4. Bannon, L. and Kuuti, K. Shifting Perspective on Organizational Memory From Storage to Active Remembering in Proceeding of the HICSS'96, IEEE Computer Press, (1996), 156-167
5. Bedard, J. Expertise and its Relation to Audit Decision Quality, Contemporary Accounting Research, Fall, pp.198-222 (1991)
6. Bennis, W., Organizing genius: the secrets of creative collaboration Addison-Wesley: Reading, Mass. (1997)
7. Berners-Lee, T., Why RDF model is different form the XML model (1998) available online: http://www.w3.org/DesignIssues/RDF-XML.html
8. Bishr, Y., Kuhn, W. Ontology-Based Modelling of Geospatial Information 3rd AGILE Conference on Geographic Information Science, Finland, May 25th-2th, (2000)
9. Bishop, K., Heads or Tales: Can Tacit Knowledge Really be Managed Proceeding of ALIA (2000) Biennial Conference, 23-26 October, Canberra, available online at http://www.alia.org.au/conferences/alia2000/proceedings/karen.bishop.html
10. Bray, T., Paoli, J., Sperberg-McQueen, C., and Maler, E., Extensible Markup Language (XML) 1.0. W3C Recommendation, 6-October-2000. http://www.w3.org/TR/REC-xml
11. Brickley, D. and Guha, R.V., Resource Description Framework (RDF) Schema Specification 1.0, W3C Candidate Recommendation, World Wide Web Consortium, (2000), http://www.w3.org/TR/rdf-schema
12. Busse, S., Kutsche, R., Leser, U., and Weber, H. Federated Information Systems: Concepts, Terminology and Architectures Forschungsberichte des Fachbereichs Informatik 99-9, (1999) available online at http://citeseer.nj.nec.com/busse99federated.html
13. Clark, J. XSL Transformations (XSLT) Version 1.0 W3C Recommendation 16 November 1999, http://www.w3.org/TR/xslt.html

14. Clark, P., Thompson, J., Holmback, H., and Duncan, L. Exploiting a Thesaurus-Based Semantic Net for Knowledge-Based Search In Proceeding 12th Conference on Innovative Applications of AI (AAAI/IAAI'00), (2000), 988-995
15. Cross, R. and Baird, L., Technology Is Not Enough: Improving Performance by Building Organizational Memory MIT Sloan Management Review Spring (2000) Vol. 41, No.3 page 69-78
16. Cui, Z., Tamma, V., and Bellifemine, F. Ontology management in enterprises BT Technology Journal Vol. 17 No 4 October (1999)
17. Davenport, T. H. and Prusak, L. Working Knowledge: how organizations manage what they know Boston, Mass, Harvard Business School Press, (1998)
18. Davies, N. J., Stewart, S. and Weeks, R, Knowledge Sharing Agents over the World Wide Web, WebNet '98, Florida, USA, November (1998)
19. Decker, S., Mitra, P., and Melnik, S., Framework for the Semantic Web: an RDF tutorial. IEEE Internet Computing, 4 (6), November/December, (2000), 68-73
20. Eppler, M.J. Making Knowledge Visible Through Intranet Knowledge Maps: Concepts, Elements, Cases System Sciences, Proceedings of the 34th Annual Hawaii International Conference, (2001), 1530 –1539
21. Fensel, D., Angele, J., Decker, S., Erdmann, M., Schnurr, H. P., Staab, S., Studer, R., and Witt, A., On2broker: Semantic-based access to information sources at the WWW in World Conference on the WWW and Internet (WebNet99), Honolulu, Hawaii, (1999)
22. Frensch, P. A. and Sternberg, R. J. Expertise and Intelligent Thinking: When is it Worse to Know Better, Sternberg, R. (ed.), Advances in the Psychology of Human Intelligence, pp.157-188 (1989)
23. Gaines, B. R. The Collective Stance in Modeling Expertise in Individuals and Organizations, available online at http://ksi.cpsc.ucalgary.ca/articles/Collective/Collective2.html
24. Garcia-Molina, H., Papakonstantinou, Y., Quass, D., Rajara-man, A., Sagiv, Y., Ullman, J., and Widom, J. The tsimmis approach to mediation: Data models and languages. In Next Generation Information Technologies and Systems (NGITS-95), Naharia, Israel, November 1995. Extended Abstract
25. Gibson, R. ed 1996 Rethinking the Future Nicholas Brealey Publishing:London
26. Heflin, J. and Hendler, J. Searching the Web with SHOE In Artificial Intelligence for Web Search. Papers from the AAAI Workshop. WS-00-01, pages 35-40. AAAI Press, (2000)
27. Horvath, J., Working with Tacit Knowledge available online at http://www-4.ibm.com/software/data/knowledge/media/tacit_knowledge.pdf
28. Hunter, J. and Lagoze, C., Combining RDF and XML Schemas to Enhance Interoperability Between Metadata Application Profiles Tenth International World Wide Web Conference, HongKong, May (2001)
29. Karvounarakis, G., Christophides, V., and Plexousakis, D., Querying Semi-structured (Meta)Data and Schemas on the Web: The case of RDF & RDFS Technical Report 269, ICS-FORTH, (2000). available at http://www.ics.forth.gr/proj/isst/RDF/rdfquerying.pdf
30. Lassila, O. and Swick, R., Resource Description Framework (RDF) Model and Syntax Specification; World Wide Web Consortium Recommendation http://www.w3.org/TR/REC-rdf-syntax/

31. Liao, M., Hinkelmann, K., Abecker, A., and Sintek, M. A Competence Knowledge Base System as Part of the Organizational Memory In: Frank Puppe (ed.) XPS-99 / 5. Deutsche Tagung Wissensbasierte Systeme, Würzburg, Springer Verlag, LNAI 1570, March (1999)
32. Marchant, G. Analogical Reasoning and Hypothesis Generation in Auditing, The Accounting Review 64, July, pp.500-513, (1989)
33. Rabarijaona, A., Dieng, R., Corby, O., and Ouaddari, R. Building and Searching XML-based Corporate Memory IEEE Intelligent Systems and their Applications, Special Issue on Knowledge Management and the Internet, May/June (2000) 56-63
34. Seligman, L. and Rosenthal, A. The Impact of XML on Databases and Data Sharing, IEEE Computer (2001)
35. Sheth, A. Changing Focus on Interoperability in Information Systems: from System, Syntax, Structure to Semantics, in Interoperating Geographic Information Systems, M. F. Goodchild, M. J. Egenhofer, R. Fegeas, and C. A. Kottman (eds.), Kluwer, (1998)
36. Stewart, T. In interview Tom Stewart on Intellectual Capital Knowledge Inc., May (1997) available online at: http://webcom/quantera/llstewart.html
37. Stuckenschmidt, H., Using OIL for Intelligent Information Integration In Proceedings of the Workshop on Applications of Ontologies and Problem-Solving Methods at ECAI (2000)
38. Vdovjak, R., and Houben, G. RDF Based Architecture for Semantic Integration of Heterogeneous Information Sources Workshop on Information Integration on the Web (2001) 51-57
39. Wellins, R. S., Byham, W. C., and Wilson, J. M. Empowered Teams: creating self-directed work groups that improve quality, productivity, and participation Jossey-Bass: San Francisco (1993)
40. Yimam, D. Expert Finding Systems for Organizations: Domain Analysis and the DEMOIR approach ECSCW 99 Beyond Knowledge Management: Management Expertise Workshop, (1999)