Expertise Matching using RDF

by Ping Liu, Jayne Curson, Peter Dew, and Richard Drew*

Information Research Institute,
University of Leeds, Leeds, LS2 9JT U.K.
pliu@comp.leeds.ac.uk, jaynec@geog.leeds.ac.uk and dew@comp.leeds.ac.uk

**Symularity Ltd*
Leeds Innovation Centre, 103 Clarendon Road, Leeds LS2 9DF U.K.
richarddrew@symularity.com

The Semantic Web aims to make data on the web machine understandable rather than only display the content. This results in all the resources being linked to each other and supports automated discovery services.

We are interested in investigating effective use of Semantic Web technology for Expertise Matching within an academic organization. This is part of a larger project to create an organizational memory for an academic institute. The importance of Expertise Matching has been noted by many researchers [Bannon96, Ackerman98, Bishop00, Cross00, Stewart97, Yimam99, Crandall98]. They realized that more emphasis should be put on the facilitation of collaboration among people through organizational memory, as the tacit knowledge and expertise held by individuals are more important than the explicit documentation. The University of Leeds Research Expertise and Publications Information System (REPIS) is a web-based research information management system that stores information about publications and research projects from a variety of different sources. A component within REPIS is an Expertise Matcher to help people locate experts in the University of Leeds and then facilitate collaboration and knowledge transferring between academia and industries. A new project, called KiMERA, aims to enhance out reach activities between the university and external organisations, and REPIS will be one of a number of existing databases to support this. In collaboration with Symularity Ltd, we are investigating the effectiveness of existing RDF tools to integrate several databases and the advantages that might accrue such as more automated brokering services.

Our belief is that the performance of expertise matching can be improved by making use of Semantic Web technologies. The obstacle of the existing broker system is that the same concept is often expressed in different terms by academia and industry. The Semantic Web will provide the technology to aid the understanding of the concept and help to break down the barriers of communication between different communities.

To understand the current state of the art we have used three research-related databases held by individual academic departments as well as by the central administration organization. We want to integrate all these databases and other semi-structured data and unstructured data (such as personal homepages and technical reports) in order to provide a single access point to the users. In our first experiment, we have integrated the REPIS database, a database of publications from the School of Computing (with sometimes conflicting data) and a database of technical reports. The tools we have used are Protégé, RDFSViz and RDFDB. Protégé is used to create and modify reusable ontologies, RDFSViz is used to visualize the ontologies represented in RDF Schema and RDFDB is used to store and query RDF data.

The experiment showed some advantages of using the RDF model:

- An RDF model greatly improves link capability. Since it is possible for separate data sources to be linked together, RDFDB provides an easy way to navigate the data warehouse. For example, all experts who have the same expertise are linked together, once you have found somebody who has a particular expertise you can easily find other people who also have expertise in the same area.
- An RDF model strongly supports extensibility. A relational database has a very static schema and it is very difficult to make any significant changes without impacting existing code. In comparison, RDF model is dynamic and it is easy to add new information.
- An RDF model provides the ability to integrate different database resources and there is no limitation on these data sources. A relational database can only store structured data. While this is not a restriction in RDFDB, structured data, semi-structured data and unstructured data can be integrated into RDFDB. In this case, REPIS database and webpages can be integrated.

However, some limitations do exist. When the user searches the RDFDB, some of the results are listed as the index of the resource and the user has to take further actions (click each of listed result items) to find more detailed information. There are also problems related to updating data. For example, when new data has been inserted in the original database, it cannot be reflected in the integrated system immediately.

For our RDF system there are two limitations existing that are not due to the RDF model but due to RDFDB itself. (1) RDFDB does not support the RDF Schema which, therefore, makes the ontology described in the Schema irrelevant and (2) RDFDB does not support part matching, so it is impossible to search in the following way: “find all the publications whose title include the word ‘visualization’”. This is limiting our ability to build practical systems.

As an increasing number of relational databases are put on the web, there is an urgent need to integrate all these operational databases. The most important practical use of the Semantic Web is to make the system understand the implicit relationships between these databases and integrate them through the semantic layer. As a result, the databases become semantically linked rather than syntactically linked which could be done in a relational database system.

More research needs to be done to support the use of the Semantic Web: firstly on the interoperability between the different schemas; secondly on how to make the links between the concepts and the terms. We also expect more powerful searching tools for RDF data and schema. The conference will enable us to learn about advances in the Semantic Web technologies and apply them to our expertise matcher and then transfer the learning to Symularity Ltd for commercial exploitation.

References

[Ackerman98]Ackerman, Mark S., Halverson, C., *Considering an Organization’s Memory* 1996; Conference on CSCW’98 pages 39-48, ACM Press, Seattle, WA, 1998

[Bannon96]Bannon, L., Kuuti, K. *Shifting Perspective on Organizational Memory From Storage to Active Remembering* in Proceeding of the HICSS’96, IEEE Computer Press, 1996, page 156-167

[Bishop00]Bishop, K. Heads or Tales: *Can Tacit Knowledge Really be Managed* Proceeding of ALIA 2000 Biennial Conference, 23-26 October, Canberra

[Crandall98]Crandall, N. F., and Wallace, M. J. *Work and Rewords in the virtual Workplace: A "New Deal" for Organizations and Employees* New York: AMACOM - American Management Association, 1998

[Cross00]Cross, R., 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

[Protégé] http://smi-web.stanford.edu/projects/protege/

[RDFDB] http://web1.guha.com/rdfdb/

[RDFSViz] http://www.dfki.uni-kl.de/frodo/RDFSViz/

[Stewart97]Stewart, T. *Tom Stewart on Intellectual Capital* Knowledge Inc., May 1997 available at http://www.brint.com/members/online/120205/stewart.html

[Yimam99]Yimam, D. *Expert Finding Systems for Organizations: Domain Analysis and the DEMOIR approach* ECSCW 99 Beyond Knowledge Management: Management Expertise Workshop, 1999