Semantically Enabled Experts Finding System – Ontologies, Reasoning Approach and Web Interface Design

Witold Abramowicz, Elżbieta Bukowska, Jakub Dzikowski, Agata Filipowska, and Monika Kaczmarek

Department of Information Systems,
Faculty of Informatics and Electronic Commerce,
Poznan University of Economics
al. Niepodleglosci 10, 61-875 Poznan, Poland
{w.abramowicz,e.bukowska,j.dzikowski,a.filipowska,m.kaczmarek}
@kie.ue.poznan.pl
http://www.kie.ue.poznan.pl

Abstract. The main goal of the eXtraSpec system is to identify experts for the needs of organizations. Therefore, the system has not only to be able to acquire and extract information from various sources, but also requires an appropriate representation of information supporting reasoning over person characteristics, as well as an appropriate Web interface allowing to create sophisticated queries in a user-friendly manner. In this paper, we address the issue of designing an ontology, reasoning mechanism as well as a Web interface for a semantic-based expert finding system.

Keywords: Expert finding system, semantic query interface design, ontology, reasoning mechanism

1 INTRODUCTION

A proper recruitment process together with knowledge on skills and expertise of employees seems to be a key to the success of an organization. More and more often organizations while searching for new employees do not only rely on their internal information sources, but they also use data available on the Internet to locate required experts. As the data available is very dispersed and of distributed nature, a need appears to support this process using IT-based solutions, e.g., information extraction and retrieval systems, especially expert finding systems. There are many research and commercial initiatives aiming at the development of expert retrieval systems that are to provide interested parties with detailed information on people's experience and skills. One of such initiatives is the Polish project eXtraSpec[1]. In order to answer users' queries on experts, the eXtraSpec system acquires and extracts information from various

[1] http://extraspec.kie.ue.poznan.pl/

Internet sources, and finally taking advantage of Semantics, reasons over person's characteristics. Within users' queries a number of criteria such as: education, experience and skills may be used and there may be different logical dependencies between the specified criteria.

There are quite a few problems that arise. In order to fulfill the previously mentioned goals, the eXtraSpec system requires an appropriate representation of information supporting reasoning over person's characteristics. Another problem concerns the simplicity and intuitive use of the interface to formulate the queries by a user. The interface that would allow to formulate complex queries would not fulfil its goal, if users would not be able to use it in practice. Therefore, our aim was to find a balance between the possibilities offered by the reasoning and querying approach followed within the system and intuitiveness of using the querying interface.

The main goal of this paper is to present the approach followed within the eXtraSpec project and discuss the underlying motivation, which led to the development of a semantic-based mechanism to retrieve experts in its current state. In order to fulfil the mentioned goals, the paper is structured as follows. First, the related work in the area is discussed. Next, the query strategies along with the resulting requirements are presented. Next, some insights into the developed ontologies and then the reasoning mechanisms, are given. Finally, remarks on the Web interface developed follow together with the interface evaluation outcomes. The paper concludes with final remarks.

2 RELATED WORK

The need to find expertise within an organization has been for long inspiration for initiatives aiming at development of a class of search engines called expert finders [Yimam, 1996]. First systems focusing on expertise identification relied on a database like structure containing a description of experts' skills (e.g., [Yimam-Seid and Kobsa, 2003]). However, these systems faced many problems regarding the gap between a generic description of expertise and fine-grained and specific queries [Kautz et al., 1996] or accuracy and validity of stored information given the static nature of a database. Thus, other systems were proposed that focused on automated discovery of up-to-date information from specific sources such as, e.g., e-mail communication [Campbell et al., 2003]. In addition, instead of focusing only on specific document types, systems that index and mine published intranet [Hawking, 2004] or Internet documents were proposed. Currently, also the Web offers many other possibilities to find information on experts. There are a number of contact management portals or Web 2.0 portals where users can look for experts, potential employees or publish their curricula in order to be found by future employers. Some examples may be: 123people[2], Sig.ma[3],

[2] `http://www.123people.com/`
[3] `http://sig.ma/`

BizWiz[4], XING[5] or LinkedIn[6]. Recently, these portals gained much popularity. However, they offer only very limited search possibilities. Some of them, e.g., 123people, aggregate information from numerous Web sources, but they are not designed for expert finding. On the other hand, portals that focus on expert searching, e.g., LinkedIn, contain only information from one source. Therefore, these existing portals do not resolve issues addressed by the eXtraSpec project.

When it comes to algorithms applied to assess whether a given person is suitable to a given task, at first, standard IR techniques to locate an expert on a given topic were applied [Ackerman et al., 2002, Krulwich and Burkey, 1996]. Here, an expertise of a person was represented in a form of a term vector and a query result was represented as a list of relevant experts. However, these approaches suffered from the well-known IR problems. In order to address these problems, the Enterprise Track at the Text Retrieval Conference (TREC) was initiated. It resulted in further advancements of the expert finding techniques and application of numerous methods such as probabilistic techniques or language analysis techniques to improve the quality of finding systems (e.g., [Balog et al., 2006, Fang and Zhai, 2007, Petkova and Croft, 2006, Serdyukov and Hiemstra, 2008]). Finally, the Semantic Web technology has been used to enrich descriptions within expert finding systems e.g., [Dorn et al., 2007a]. The introduction of Semantics into search systems takes mainly two forms: analysis of indexed documents or queries, or operating on semantically described resources (e.g., RDF files) with a use of reasoners. In consequence also as a number of different ontologies used to represent competencies and skills were developed [Aleman-Meza et al., 2007, Gómez-Pérez et al., 2007, Dorn et al., 2007b].

The expert finding systems have interfaces similar to the regular search engines. Search engine interfaces are to help users in definition of their information needs in a form of queries as well as in the understanding, restructuring and refining the search results. However, even such simple interfaces may pose severe problems, when it comes to keyword specification. [Hargittai, 2004, Muramatsu and Pratt, 2001, Shneiderman et al., 1997] specified eight design guidelines for development of search user interfaces. These guidelines were taken into account while developing the interface for the eXtraSpec system as a user interface is often one of the most critical factors when it comes to the success or failure of a computerized system [van Vliet, 2008]. A user judges the quality of a system based on the interface and the way it helps him to accomplish the tasks. In a technical sense, we perceive user interface as an architecture layer separated from the application logic.

3 QUERYING STRATEGIES

In order to identify the requirements towards the eXtraSpec system, first some search scenarios were considered. This allowed to specify requirements towards

[4] `http://www.bizwiz.com`
[5] `http://www.xing.com`
[6] `http://linkedin.com`

the developed ontology, reasoning mechanism and GUI. The scenarios have been specified based on the conducted studies of the literature and interviews with employers. The six most common searching goals are as follows:

1. To find an expert with some experience on a position of interest. Thus, the system must be able to analyse positions and jobs undertaken so far as well as their duration. The requirement on the GUI includes enabling specification of these criteria, i.e., to point to a job name (a position of interest) and to add the length of a required experience. The requirements on the reasoning mechanism include ability to integrate the history of experience (e.g., add the length of duration from different places but gained on the same or similar position) and the ability to reason on a positions' hierarchy (i.e., taking into account narrower or broader concepts). This implies also a requirement on the underlying ontology to represent a is-a hierarchy of different positions and jobs.

2. To find an expert having some specific language skills – Features of interest for this scenario include: known languages, obtained certificates and a level of language skills. Regarding requirements for GUI, there is a need to point to a language of interest (a list), indicate the proficiency level of the language and (if needed) a name of the certificate. If the information is not explicitly given, the reasoner needs to be able to associate different certificates with languages and proficiency levels. Ontologies must represent languages certificates (is-a hierarchy) together with information on the language and the proficiency level, mapped to one scale.

3. To find an expert having some competencies – The reasoning mechanism is to operate on indexed competencies and skills. The GUI has to provide a possibility to point to a name of skills/competencies of interest. A reasoner should be able to operate not only on explicitly given competencies but also reason on jobs and then on connected competencies. Thus, the reasoner needs to tackle also other relations than is-a. Ontologies have to include skills and competencies hierarchy and some additional relations.

4. To find students who graduated recently/will graduate in a given domain – In this scenario the important features are: educational organization, date of graduation and educational result. The requirement for GUI is to point to a category of educational organization or a specific organization, to name the result, start and end date of education. The reasoning mechanism has to reason on the hierarchy of educational organizations, on dates and results. Regarding the ontology, the considered scenario requires a hierarchy of educational organizations.

5. To find a person having expertise in a specific domain – This scenario operates on a list of organizations and on domains they operate in, thus, this data has to be provided by the ontology. The requirement on GUI is to specify a domain of interest. The reasoner has to be able to associate organizations with domains.

6. To find a person with specific education, competencies, jobs, etc. – Features of interests include all previously mentioned. The requirement on GUI is to give possibility to define numerous queries with all previously specified goals and combine various categories into a one complex query and make it as easy as

possible to specify various logically connected constraints. The reasoner has to combine results from various categories using different logical operators. The requirements on ontologies are the same as in scenarios 1-5.

As already discussed, the above querying strategies imposed some requirements on the information on experts that should be available as well as ontologies that needed to be developed for the project needs.

4 ONTOLOGIES

The ontology underlying the system needs to support not only the scenarios mentioned above. The eXtraSpec system by acquiring automatically data from the dedicated Web sources, imposes some additional requirements on the ontology. Namely, the developed ontology should enable the semantic annotation of all elements of experts' profiles as well as support the normalization and discovery process. Each expert is described with series of information, e.g., first and last name, history of education, career history, hobby, skills, obtained certificates. To make the reasoning possible, the following attributes from the profile of an expert should be linked to ontology instances: educational organization; certifying organization; client, employer and fulfilled role; scope of education; topic of education; result of education; skill name; certificate name and finally degree of a skill. For more details see [Abramowicz et al., 2011b].

Performed analysis of the requirements imposed on the ontology, concluded with the definition of a set of relations that should be defined:

- subConceptOf representing hierarchical relations between concepts;
- isPartOf representing composition of elements;
- isRequired representing a connection between two concepts;
- implies from one fact, or set of facts, another fact can be concluded.

We have used the OWL language as the underlying formalisms of the developed ontology and the SKOS model as a data model. The data structure was designed having one SKOS ontology with eight concept schemas for each area of interest: Organizations (for organizational organizations, certifying organizations, Employer and Client), SkillName, SkillDegree, Certificate, Role, EducationScope, EducationTopic and EducationResult. While building the ontology for the needs of the eXtraSpec system, a wide range of taxonomies and classifications has been analyzed in order to identify the best practices and effective solutions. As the eXtraSpec system is a solution designed for the Polish market, so is the developed ontology.

5 REASONING MECHANISM

In order to support the searching scenarios, the eXtraSpec system needs not only an appropriate representation of information on person's characteristics, but also the reasoning mechanism being efficient and scalable. In addition to the requirements already mentioned, the following requirements for the querying and reasoning mechanism also need to be considered:

- building queries in a structured way (i.e., feature: desired value);
- supporting definition of desired values of attributes in a way suitable to the type of data stored within the given feature;
- joining a subset of selected criteria within the same category into one complex requirement using different logical operators;
- formulating a set of complex requirements within one category with different logical operators;
- joining complex requirements formulated in various profile categories into one criteria with different logical operators;

The logical operators between different criteria and their values include such operators as: must, should, must not.

The eXtraSpec system consists of a number of modules specialized to carry out different tasks (see [Abramowicz et al., 2010] for details). One of the components is the reasoning component REA consisting of the following mechanisms: indexing mechanism (indexer), searching mechanism (searcher), composition mechanism (composer) and a set of ontologies, together with the reasoning engine. REA supports two independent processes. First, creating indexes of profiles — optimized for search, i.e., structured so as to enable a very fast search based on criteria pre-set by a user. The profile is analysed, divided into relevant sections, and then enriched with additional information using an ontology (pre-reasoning). The second process that needs to be supported is defining the query matching mechanism on the enriched indexes - this process is initiated by the task of a user formulating queries using a graphical interface that is presented and discussed in details within the next section.

To support the IR side of the mechanism, the open-source java library Lucene[7], supported by the Apache Software Foundation, was selected. However, fields in the Lucene documents cannot be grouped together nor stored as hierarchical structures what is needed in the light of the defined requirement. Thus, during the indexing process profiles are divided into a number of separate documents. Concurrently with the indexing process, pre-reasoning takes place, in order to complete profile with the implied facts (see [Abramowicz et al., 2011b] for details).

6 EXTRASPEC WEB INTERFACE DESIGN AND EVALUATION

The front-end to the eXtraSpec system should enable users to build complex queries describing characteristics of desired experts. The system must enable a user to specify constraints on expert's attributes and select whether the value of an attribute is required, desired (but not required) or not allowed. In addition, the interface should: enable for grouping of constraints; provide a possibility to build queries including complementary and alternative constraints; enable providing some of criteria values typed as free text and some of them to be

[7] http://lucene.apache.org

chosen from the eXtraSpec system knowledge base; be loosely coupled with the system; and finally be understandable and easy to use. Besides, the complexity of querying should not affect the interface usability. An average computer-skilled user should facilely express his or her information need.

The conceptual model of the interface is determined by the scheme of querying the experts finding system and the structure of the profile. The search criteria are divided into the following categories: personal data, education, professional experience, foreign languages, courses, certificates, additional skills, organization membership and interests. Each category consists of a group of fields. Desired values of these fields are specified in the interface by criteria values, and field groups by criteria groups. Each criterion has a label and a value typed by the user, selected from list or from a tree of possible values.

A user can add a criteria group by clicking a button "Add criteria" or selecting a group name from the list. A criteria group can include one or more criterion. Within the group user may specify alternate criteria by clicking a button "or..." or "neither..." (depending on the type of constraint for the group). A user may also type some of the criteria, for example a last name or an e-mail address; with the use of wildcards, when needed. Besides, some criteria values depend on the knowledge base of the eXtraSpec system, and possible values are loaded from the ontology. A user may specify them by clicking on a particular field and then select appropriate values from the tree in a pop-up window. The selected items are presented on the bottom of the pop-up window and, finally, in the appropriate field.

The experiment in order to evaluate the quality of the interface was carried out in two distinct phases with a 2-weeks' break in-between. In both phases the same group of user's participated. The group consisted of 31 students in their final year of the bachelor studies.

The first phase was focused on learning how to use the interface and pose the queries using the system. The users were provided with a set of free text queries (considering query strategies presented in section 3) and asked to introduce these queries into the system using the provided interface. All users could choose how many queries they introduce into the system (on average they included 2.2 query). As a result 70 query instances with an average accuracy at 83% were gathered (all generated queries were checked for completness). The best accuracy (95%, 12 query instances) users reached for query "Computer scientists with at least 3-years of experience, required English, desired German", worst (56%, 8 instances) for "Candidates who graduated in the last 3 years from a university except from universities of economics, in the field of computer science". This phase of the experiment showed that the current interface makes the system easy to use, but the level of the error tolerance or user assistance while generating queries should be improved. The users also knew that they may influence the final version of the interface, so they provided also recommendations what should be changed within the system.

In the second phase users again were asked to experiment with the system, however this time they were asked to provide detailed marks for each interface

feature. The aim of this phase of the experiment was to evaluate the interface quality. Therefore, users were to answer 12 detailed questions about the system features. The best rating the application front-end gained in the capabilities category and the worst in the ease of use category (mainly because of not enough support offered to the user while defining the query).

The most interesting however are not the average evaluation results for each category, but their distribution when it comes to different values from the scale. The system was judged as good or very good by most of the users. They also underlined that the system is suitable also for a user without the expertise in IT (especially when it comes to the advanced querying). Users also evaluated the clarity and distribution of information on the interface. The eXtraSpec interface being much more complex when compared to the Google search engine interface, is still understandable. Moreover, the quality of search results compensates the querying effort. For detailed results of the evaluation please see [Abramowicz et al., 2011a].

7 CONCLUSIONS

The main goal of the eXtraSpec project is to develop a system supporting analysis of company documents and selected Internet sources for the needs of searching for experts from a given field or with specific competencies. The provided system focuses on processing texts written in the Polish language. The obtained information is stored in the system in the form of experts profiles and may be consolidated when needed. The system offers also a user friendly interface to perform queries that allow to find persons with specific characteristics. The realisation of this goal required interconnection between developed interface and underlying ontologies.

Within this paper, we have discussed the concept and considered scenarios regarding the implementation of the reasoning mechanism for the needs of the eXtraSpec system together with its querying interface allowing to find persons with specific characteristics. The conducted evaluation has proven that extending the query definition process with a form that need to be filled in before submitting the query, does not pose problems to users and quality of results recompensates the effort needed for query definition.

The set of developed ontologies discussed within this paper was designed specially for the Polish language, however, the main structure and model as well as defined relations may be reused also for other languages.

The future work focuses on implementing the system in companies from the construction sector and further interviewing the system users on the level of usefulness and interface quality when it comes to the query specification and fullfilment of the information needs. This is also related with analysis of improvements offered by a semantically-enabled search in comparison to a typical database-driven approach.

ACKNOWLEDGEMENTS

The work published in this article was supported by the project titled: "Advanced data extraction methods for the needs of expert search" financed under the Innovative Economy Framework and partially supported by European Union in the European Regional Development Fund programme (contract no. UDA-POIG.01.03.01-30-150/08-01).

References

[Abramowicz et al., 2011a] Abramowicz, W., Bukowska, E., Dzikowski, J., Filipowska, A., and Kaczmarek, M. (2011a). Web interface for semantically enabled experts finding system. In *Proceedings of the 13th International Conference on Enterprise Information Systems*.

[Abramowicz et al., 2011b] Abramowicz, W., Bukowska, E., Kaczmarek, M., and Starzecka, M. (2011b). Semantic-enabled efficient and scalable retrieval of experts. In *International Conference on Information, Process, and Knowledge Management, eKNOW 2011*.

[Abramowicz et al., 2010] Abramowicz, W., Kaczmarek, T., Stolarski, P., Wecel, K., and Wieloch, K. (2010). Architektura systemu wyszukiwania ekspertów eXtraSpec. In *Technologie Wiedzy w Zarzadzaniu Publicznym*.

[Ackerman et al., 2002] Ackerman, M., Wulf, V., and Pipek, V. (2002). *Sharing Expertise: Beyond Knowledge Management*. MIT Press.

[Aleman-Meza et al., 2007] Aleman-Meza, B., Bojars, U., Boley, H., Breslin, J., Mochol, M., Nixon, L., Polleres, A., and Zhdanova, A. (2007). Combining rdf vocabularies for expert finding. In *Proceedings of the 4th European Semantic Web Conference (EWSC 2007)*.

[Balog et al., 2006] Balog, K., L., A., and De. Rijke, M. (2006). Formal models for expert finding in enterprise corpora. In *Proceedings of the ACM SIGIR*, pages 43–50.

[Campbell et al., 2003] Campbell, C. S., Maglio, P. P., Cozzi, A., and Dom, B. (2003). Expertise identification using email communications. In *CIKM '03: Proceedings of the twelfth international conference on Information and knowledge management*, pages 528–321. ACM Press.

[Dorn et al., 2007a] Dorn, J., Naz, T., and Pichlmair, M. (2007a). Ontology development for human resource management. In *Proceedings of 4th International Conference on Knowledge Management*, Series on Information and Knowledge Management, pages 109–120.

[Dorn et al., 2007b] Dorn, J., Naz, T., and Pichlmair, M. (2007b). Ontology development for human resource management. In *Knowledge management: innovation, technology and cultures: proceedings of the 2007 International Conference on Knowledge Management, Vienna, Austria, 27-28 August 2007*, volume 6, page 109. World Scientific Pub Co Inc.

[Fang and Zhai, 2007] Fang, H. and Zhai, C. (2007). Probabilistic models for expert finding. In *Proceedingsof the ECIR*, pages 418–430.

[Gómez-Pérez et al., 2007] Gómez-Pérez, A., Ramírez, J., and Villazón-Terrazas, B. (2007). An ontology for modelling human resources management based on standards. In *Proceedings of the 11th international conference, KES 2007 and XVII Italian workshop on neural networks conference on Knowledge-based intelligent information and engineering systems: Part I*, pages 534–541. Springer-Verlag.

[Hargittai, 2004] Hargittai, E. (2004). Do you "Google"? Understanding search engine use beyond the hype. *First Monday*, 9(3):01.

[Hawking, 2004] Hawking, D. (2004). Challenges in enterprise search. In *Proceedings of the 15th Australasian database conference – Volume 27*, ADC '04, pages 15–24, Darlinghurst, Australia, Australia. Australian Computer Society, Inc.

[Kautz et al., 1996] Kautz, H., Selman, B., and Milewski, A. (1996). Agent amplified communication. In *Proceedings of the Thirteenth National Conference on Artificial Intelligence (AAAI-96)*, pages 3–9.

[Krulwich and Burkey, 1996] Krulwich, B. and Burkey, C. (1996). Contactfinder agent: answering bulletin board questions with referrals. In *Proceedings of the National Conference on Artificial Intelligence*, pages 10–15.

[Muramatsu and Pratt, 2001] Muramatsu, J. and Pratt, W. (2001). Transparent queries: investigation users' mental models of search engines. In *Proceedings of the 24th annual international ACM SIGIR conference on Research and development in information retrieval*, pages 217–224. ACM.

[Petkova and Croft, 2006] Petkova, D. and Croft, W. (2006). Hierarchical language models for expert finding in enterprise corpora. In *Proceedings of the 18th IEEE International Conference on Tools with Artificial Intel-ligence*, pages 599–608.

[Serdyukov and Hiemstra, 2008] Serdyukov, P. and Hiemstra, D. (2008). Modeling documents as mixtures of persons for expert finding. In *Proceedings of the ECIR*, pages 309–320.

[Shneiderman et al., 1997] Shneiderman, B., Byrd, D., and Croft, W. (1997). Clarifying search: A user-interface framework for text searches. *D-lib magazine*, 3(1):18–20.

[van Vliet, 2008] van Vliet, H. (2008). *Software Engineering*. John Wiley & Sons, Ltd.

[Yimam, 1996] Yimam, D. (1996). Expert finding systems for organizations: Domain analysis and the demoir approach. In *ECSCW 999 Workshop: Beyond KNowledge Management: Managing Expertise,*, pages 276–283, New York, NY, USA. ACM Press.

[Yimam-Seid and Kobsa, 2003] Yimam-Seid, D. and Kobsa, A. (2003). Expert finding systems for organizations: Problem and domain analysis and the demoir approach. In *Journal of Organizational Computing and Electronic Commerce*, 13(1), pages 1–24.