

Suitable employees wanted? Find them with semantic techniques

Malgorzata Mochol¹, Anja Jentzsch¹, Holger Wache²

¹ Freie Universität Berlin, Institut für Informatik, Takustr. 9, 14195 Berlin, Germany
E-mail: mochol@inf.fu-berlin.de, anja@anjeve.de

² University of Applied Sciences Northwestern Switzerland (FHNW)
Riggenbachstrasse 16, 4600 Olten, Switzerland, E-mail: holger.wache@fhnw.ch

Abstract. More and more companies are going to use Internet for their job procurement process, i.e. to search for job seekers and publish open positions. However, common search techniques based on classical keyword search fail to match job seekers with job offers since particular (over-constraint) queries return no answers and search engines are unable to semantically relax such queries. In this paper we briefly describe an approach where explicitly defined rewriting rules are used to reformulate the query in a controlled way and explain the architecture of our job portal extension giving an impression about the relaxation process.

1 Introduction

The Internet gains continuously in importance regarding job procurement process since more than three-quarter of the age class, which is relevant for the recruitment, is online, and due to the fact that increasingly more companies publish their job openings on the web. There is a large number of online commercial job portals competing to publish job postings for a fee, on the other hand, each company can publish job postings on its company's own Website [4]. However, publishing postings on the corporate Website reaches a very limited audience, because the indexing capabilities of current search engines are too imprecise to support searches for open positions. Beside this, meta-search engines are limited in their ability to generate offers that match the precise needs of the clients since job postings are written in free text form using uncontrolled vocabulary. Furthermore, some dedicated search engines are entering into the market, allowing detailed queries as opposed to keyword-based search of current search engines. However, the quality of search results depends not only on the search and index methods applied. Influential factors include the processability of the used Web technologies and the quality of the automated interpretation of the company-specific terms occurring in the job descriptions. The problems of a website's machine processability result from the inability of current web technologies to semantically annotate the content of a given website. This means that from the perspective of the Semantic Web there are capabilities for improvement for all parties involved in the recruitment scenario based on the semantic annotation of

job postings and profiles. In our opinion, using semantic technologies in the domain of e-Recruitment and skill management can substantially increase market transparency, lower the transaction costs and speed up the procurement process. In this work we are reporting on the results of a cooperation between the German project *Knowledge Nets*³ that has been working together with one of the job portal and *European Network of Excellence Knowledge Web*⁴. The paper is organized as follows: Sec. 2 gives an overview of the prototypical Knowledge Nets-job portal that is based on semantic matching techniques. Having identified the importance of e-Recruitment processes for businesses, we have developed an innovative approach to online recruitment based on the semantic technologies from the first prototype in combination with query approximation (Sec. 3) and realized this idea in the prototypical implementation described in Sec. 4. We summarize our work with a brief conclusion in Sec. 5.

2 Semantic job search

The project Knowledge Nets approaches the impact of semantic technologies from the business and technical viewpoint in order to make predictions about the influence of these new technologies on markets, enterprises and individuals. For this purpose the project takes a closer look at particular market sectors and application scenarios. The first scenario situated in the Human Resource (HR) domain aims to analyze the online job seeking and procurement processes with and without the usage of semantic technologies [2, 5, 11]. In a Semantic Web-based recruitment application the data exchange between employers, job applicants and job portals is based on a set of shared vocabularies describing domain relevant terms: occupations, industrial sectors, skills and educational background. This means that the first step towards the realization of the semantic-based e-Recruitment scenario was the creation of a human resources ontology (HR ontology) that describes formally these commonly used vocabularies⁵. The requirements analysis revealed the necessity of aligning the resulting ontology with commonly used domain standards and classifications in order to maximize the integration of job seeker profiles and job postings.

Given a rich and machine-processable representation of the domain of interest, the ontology forms the basis for the implementation of semantic matching which combines annotations using controlled vocabularies with background knowledge about a certain application domain. In our prototypical implementation [8], the domain specific knowledge is represented by concept hierarchies like skills, occupational classification, and a taxonomy of industrial sectors. Having this background knowledge of the recruitment domain allows us to compare job descriptions and applicant profiles based on their semantic similarity [10] instead

³ <http://wissensnetze.ag-nbi.de>

⁴ <http://knowledgeweb.semanticweb.org>

⁵ The reuse-oriented methodology for ontology engineering that has been applied for the creation of the HR ontology is described in [9] and our experiences in the building of e-Recruitment applications are stated in [6].

of merely relying on the containment of keywords like most of the contemporary search engines do. The result of semantic matching is a ranking list of the best fits of openings for the job seeker and likewise ranking list of job applicants for the employer. In the detail description of each recommended job offer or applicant profile there is an explanation why an opening fits an applicant. Explanations of this kind is absent in current job search portals and websites. The application of ontologies and the semantic matching approach contribute to the realization of more powerful and flexible e-Recruitment solutions which include advanced search, crawl and presentation facilities based on knowledge about the application domain. However, in the course of further development and testing of our system we noticed that the mentioned above solution alone can not meet real world requirements, which include the possibility of overly specific or inconsistent queries and that there are still some issues open.

3 Query approximation in Human Resource

Usually the description of the job requirements are too specific to be matched, and in turn, delivering only a few or even zero answers immediately disappoints users of job portals. To be able to tackle such issues and to satisfy the real world requirements we introduce the rewriting rules approach to capture the domain and user-specific knowledge explicitly, i.e. we apply query relaxing with rewriting rules that, in turn, delivers a set of rewritten queries. The results of each rewritten query may be, however, a set of answers which is not ordered. For the ordering purposes we can fall back to the semantic similarity function with their implicitly encoded knowledge and rank the answers for each rewritten query. The query rewriting approach provides a high level relaxation including grouping the results according the domain knowledge and the user preferences whereat the similarity function provides some kind of fine tuning when the results in one group is ranked. The rewriting relaxes the over-constraint query based on rules in order defined by conditions. We start with the strongest possible query that is supposed to return the “best” answers satisfying most of the conditions. If the returned result set is either empty or contains unsatisfactory (number of) results, the query is modified either by replacing or deleting further parts of the query, i.e. the *query is relaxed* ⁶.

Each job request and each opening is annotated with an RDF description which is a set of triples. A query over these job openings is formulated as triple patterns and a set of conditions that restrict the possible variables bindings in the patterns. Each triple pattern represents a set of triples and the corresponding abstract definition of a query focuses on the essential features of queries over RDF. We can use generic rewriting rules that are guided by information from the ontology and specify the relaxation by making use of the predicates e.g. by exchanging the required experience X with Y when X is more specific than Y:

```
IF experience = X  
REPLACE-BY experience = Y
```

⁶ We have utilized the rule-based query rewriting framework for RDF queries [3].

WHEN Y is more general than X

Following the general rewriting rule a specific rule taking into account a required skill, educational background like academic degree or certain experiences can be defined. In a particular case this means that considering e.g. the required experiences, which are annotated using the skill ontology⁷, we can replace concept `JAVA` that is specified in the original query by the concept `ObjectOrientedLanguages` since the latter is more general than the first. Other way to relax the query is the application of such rewriting to number restrictions. If we consider an opening that requires from job seeker to have 5 years project experience, we can relax the query by exchanging 5 with 3 years. In the following section we will briefly present the first prototypical implementation of our e-Recruitment portal which is based on the query relaxation technique.

4 HR Prototype

In terms of the cooperation between the Knowledge Nets project and the Knowledge Web network we have been developing a prototype of a job portal that (i) takes into account the real world requirements derived from the industrial use cases collected in the framework of Knowledge Web [1, 7], (ii) utilizes the HR ontology developed in Knowledge Nets, (iii) applies the rewriting rules as a simple technique for query relaxation and (iv) uses the semantic matching approach to rank the final results. The HR prototype consists of three main layers: information consumer, information aggregator and information provider (cf. Fig. 1).

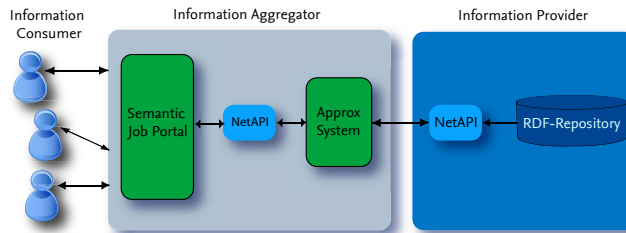


Fig. 1. Architecture of the HR Prototype

Information consumers, the companies with open position (or job seekers), specify their needs (or profiles) concerning i.e. required experiences in particular technologies or in certain industrial sectors, the necessary skills, the educational background and the certain occupations of potential applicants (or its open job positions) using the web interface offered by information aggregator .

From the technical point of view the requirements defined by the employers are converted to a SeRQL query and forwarded to the *Approx System*. *Approx System* based on SWI-Prolog⁸ is responsible for the relaxation of the particular queries as described in Sec. 3. and maintain the rewriting rules. Since *Approx*

⁷ The skill ontology is a part of our HR ontology

⁸ <http://www.swi-prolog.org/>

has not only to apply the rewriting rules to the given query but, in the first step, it must detect suitable rules, it consists of two main components: *controller* that controls the search process and the *rewriter* which is responsible for applying the rewriting rules to a query. Given a query, the controller asks the rewriter for all possible relaxation of this query. The returned relaxed queries are checked by the controller if they return answers. The controller uses for this purpose its interface the RDF Repository Sesame⁹ where the developed HR ontology together with the instance data regarding candidate profiles are stored. All queries which do not return any answers are maintained in a list by the controller and are further relaxed with the help of the rewriter until the rewriter can not relax the query any more or the relaxed query returns at least one answer.

Let's consider a simple example where a company is looking for a suitable applicant with the following requirements: over 10 years experience in national projects with emphasis on Research & Development in the area of natural sciences and mathematics and some competencies in XML.

Wissensnetze Recruitment Scenario

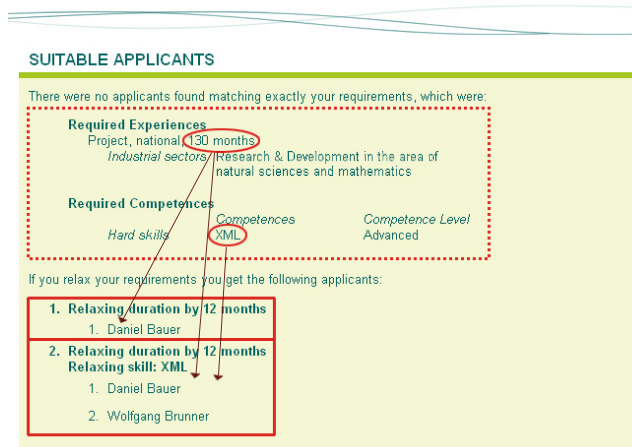


Fig. 2. Semantic Job Portal: suitable applicants after relaxation

After specification of the requirements regarding the potential candidate using the web interface of the Semantic Job Portal, the system is looking for suitable profiles of job seekers stored in the system. Due to the fact that there are no suitable candidates matching the primary requirements the controller from the Approx System will choose i.e. the following rewritings (cf. Fig. 2): (1) relax the length of the experiences: 9 years experience instead of more than 10, and (2) - relax the length of the experiences and additionally the hard skills by removing from the original query the requirements containing XML. After the Approx System rewriter applies the first rule, the system, as shown in Fig. 2, delivers one suitable candidate, Daniel Bauer, and after deploying of the second rule there are two persons who match the query, Daniel Bauer and Wolfgang Brunner.

⁹ <http://www.sesame.org/>

5 Conclusion

The idea of rewriting rules is quite simple and not really new but just in the combination with semantic technologies it has evolved to a powerful approach and, in the context of e-Recruitment, proved to be a valuable and helpful technique. The query approximation is useful in areas where too specific queries need to be loosened in order to allow users to find best matches rather than simply receive no results at all. On top of the presented prototype we plan to apply the semantic matching techniques in order to rank the results provided by the approach of query relaxation.

Acknowledgement The work described in this paper is supported by the EU Network of Excellence KnowledgeWeb (FP6-507482) and Knowledge Nets project which is a part of the InterVal- Berlin Research Centre for the Internet Economy funded by the German Ministry of Research (BMBF).

References

1. L. Nixon A. Leger, F. Paulus and P. Shvaiko. Towards a successful transfer of knowledge-based technology to European Industry. In *Proc. of the 1st Workshop on Formal Ontologies Meet Industry (FOMI 2005)*, 2005.
2. C. Bizer, R. Heese, M. Mochol, R. Oldakowski, R. Tolksdorf, and R. Eckstein. The Impact of Semantic Web Technologies on Job Recruitment Processes. In *International Conference Wirtschaftsinformatik (WI'05)*, 2005.
3. P. Dolog, H. Stuckenschmidt, and H. Wache. Robust query processing for personalized information access on the semantic web. In *7th International Conference on Flexible Query Answering Systems (FQAS 2006)*. Springer, June 2006.
4. W. Mülder. Personalinformationssysteme - Entwicklungsstand, Funktionalität und Trends. Human Resource Information Systems. *WIRTSCHAFTSINFORMATIK. Sonderheft IT & Personal*, 42:98–106, 2000.
5. M. Mochol, R. Oldakowski, and R. Heese. Ontology-based Recruitment Process. <http://page.mi.fu-berlin.de/mochol/papers/SemTech.pdf>, 2004.
6. M. Mochol and E. Paslaru Bontas. Practical Guidelines for Building Semantic eRecruitment Applications. In *International Conference on Knowledge Management, Special Track: Advanced Semantic Technologies (AST' 06)*, 2006.
7. L. Nixon and M. Mochol. Prototypical business use cases. Technical report, Deliverable 1.1.2 in the Knowledge Web EU Network of Excellence, 2005.
8. R. Oldakowski and C. Bizer. SemMF: A Framework for Calculating Semantic Similarity of Objects Represented as RDF Graphs. In *Poster at the 4th International Semantic Web Conference (ISWC 2005)*, 2005.
9. E. Paslaru Bontas and M. Mochol. Towards a reuse-oriented methodology for ontology engineering. In *Proc. of 7th International Conference on Terminology and Knowledge Engineering (TKE 2005)*, 2005.
10. J. Poole and J.A. Campbell. A Novel Algorithm for Matching Conceptual and Related Graphs. *Conceptual Structures: Applications, Implementation and Theory*, 954:293–307, 1995.
11. R. Tolksdorf, M. Mochol, R. Heese, R. Eckstein, R. Oldakowski, and C. Bizer. Semantic-Web-Technologien im Arbeitsvermittlungsprozess. *Wirtschaftsinformatik: Internetökonomie*, 48(1):17–26, 2006.