Use Case 1 in Service Industry – Research Challenges
Recruitment

KW Partner: FU Berlin
IB Member: World Wide Jobs GmbH

1. General Description of Business Use Case

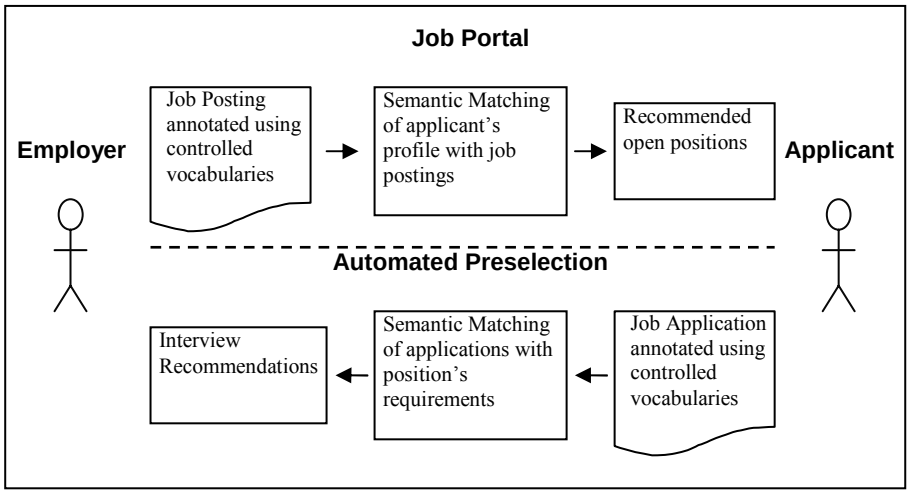
The recruitment of employees is an important practice for any business and it is increasingly being carried out online. For example, in 2004 almost 28 % of 14-64 years old people used the web to research the employment market and look for open job positions[1]. In the near future over 50% of recruitments made in Germany are expected to be the result of an online job posting. Finding the best suited candidate in the shortest time leads to cost cutting and resource sparing by the Human Resources department and can potentially bring firms into contact with a higher level of quality of candidates.

The aim of an online recruitment system is to facilitate filling open job vacancies with qualified and suitable candidates. There are many different portals and websites with open positions nowadays that the job seeker can in no way gain an overview of all the offers published on the internet.

On the other side more and more companies post their vacancies only on their company websites. The problem is that most of these sites cannot be found by the popular search engines due to the lack of "meaning" (semantic), i.e. they are not recognised as "job postings". This results in a lower number of applicants and a longer time necessary for hiring. The other problem is the low level of matching between candidates and vacancies, which is based on string comparison and simple categorization. As a result a significant amount of manual effort is still required to find a appropriate job offer or sort through applications in order to even filter out applications that to a human observer are clearly irrelevant.

Equally, assessment of candidate suitability to a vacancy is a manual task, with the computer system unable to provide any means for an initial ranking of applicants. The potential of Semantic Web technologies to match between concepts in a description (vacancies and candidates) with improved quality of query results and to generate rankings of matches is relevant to this business problem. Moreover, the identification of vacancies and their machine-interpretability are in this case highly essential tasks.

[1] ACTA – Allensbacher Computer and Technique Analysis; http://www.acta-online.de/

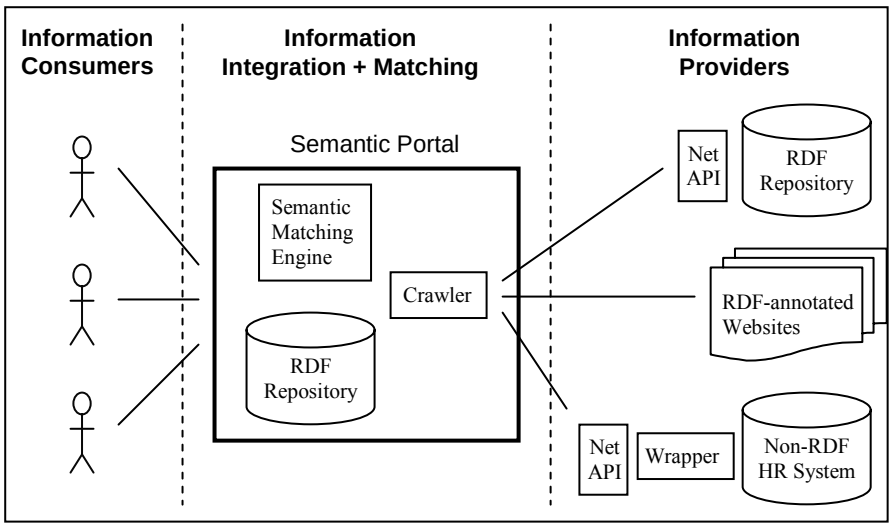


2. Proposed Semantic Web-based Solution

In order to evaluate a knowledge-based solution, a semantic job portal is being prototyped. The system is expected to fulfill the following requirements:

- Easy to use: a user can intuitively interact with the system without any prior knowledge about the design of ontologies;
- High quality of query results: by means of a semantic matching algorithm the number of false positives and false negatives is minimized, and results are ranked;
- Flexible matching: the user has some control on the similarity measure of the matching algorithm;
- Efficiency: the matching algorithm should operate efficiently;
- Configurability: the components of the system should be changeable with minimal effort, i.e. modification of the matching algorithm.

A general overview of the proposed solution shows the participants and their roles.



Steps already taken in this scenario are:

- The development of a Human Resources ontology, created by the modification and integration of existing standards such as the HR-XML vocabulary and various classification schemes (e.g. occupation and industry classification);
- The implementation of a prototypical system, which consists of an user interface front-end (JSP), RDF storage and an engine for extracting the knowledge stored in the database;
- Trials on matchmaking solutions, which should match between job applicants and job postings and return a ranked results list. The current solution focuses on taxonomical and attributive similarity that depends on the presence of particular concepts, the distance between them and relations.

In the next step the scalability of this prototypical implementation will be tested with larger volume of data. Later it is planned to integrate other job information sources through a crawler. This requires adding support for the alignment or merging of other ontologies as well as using wrappers to extract knowledge from non-RDF systems.

3. Identified Research Challenges

3.1 Semantic Matching

3.1.1 Problem Statement

Inside both job postings and job applications sub-ontological pieces of information can be grouped into "thematic clusters", e.g. information about competencies and skills, information regarding the industry sector of the job position, and job position details. The thematic cluster from a job posting is compared with the corresponding cluster from a job application. General similarities are calculated from an average of cluster similarities which result from the semantic similarity between concepts. We measure taxonomical similarity through the distance between the concepts, i.e. their respective positions in the concept hierarchy and the attributive similarity which is based on the comparison of attributes and their values. However some limitations to this approach become apparent.

For example, we noticed that, unlike "hard skills" (certifiable such as language or qualification), "soft skills" (uncertifiable such as creativity or teamwork skills) proof to be not as easy to quantify and hence are more difficult to automatically compare. Due to this limitation the last phase of recruitment process (the end decision) still has to be done by humans and cannot be automated even with the use of Semantic Web technologies.

Matchings may also be made more exact by adding weightings to concepts (i.e. describing or measuring the importance of concepts) or associating them with competency levels.

3.1.2 Knowledge Processing Task and Component

The identified knowledge processing tasks are matching and ranking. Matching refers to the task of discovering relationships between entities in ontologies and measuring the level of similarity between two entities. Ranking refers to ordering match results according to a desired criterion. In this use case, as exact matches between job requirements and applicants are unlikely to happen, the ranking mechanism is used to express the extent to which equivalence might be assumed. Both tasks are considered as occurring in the Semantic Matching Engine. The engine provides a similarity relation between two concepts in the form of a coefficient in the [0,1] range and uses this as the basis for results ranking.

3.1.3 Requirements Analysis

The requirements upon the Semantic Matching Engine, i.e. upon the matchmaking algorithm that it uses, are:

- Support the weighting of concepts as a means of tweaking matching results;
- Support a more precise matching between job position postings and seekers and better ranking of results (i.e. extending the criteria from solely using the similarity coefficient), including the facility to provide natural language explanations of the ranking;
- Enable different rankings with respect to the thematic clusters defined in the ontology
- Support the consideration of measures in the matchings, such as competency levels (e.g. novice, intermediate, expert) or duration (e.g. number of years in a job).
- Support of parameterisation of matching algorithm which allows the matching of other ontologies in other scenarios.

3.2 Storage and Retrieval

3.2.1 Problem Statement

Such a portal will have to store huge amounts of collected RDF data and provide high performance access to this data for its users. This raises the problem of scalability and efficient retrieval in current RDF storage solutions.

3.2.2 Knowledge Processing Task and Component

The storage of instance data will be handled by the directory manager, who provides the necessary functionality to add, update, remove or query RDF triples.

3.2.3 Requirements Analysis

An underlying storage implementation is required that supports the scalable storage of a large volume of RDF triples. Through the directory manager component, operations on this large scale repository need to demonstrate high performance and reliability.

3.3 Knowledge Extraction

3.3.1 Problem Statement

Employers want to publish their job postings through their existing software infrastructure. Hence means are needed to generate RDF data using the HR ontology from legacy data sources which could be flat text files, HTML or tables in relational databases.

3.3.2 Knowledge Processing Task and Component

Two approaches are possible: if the legacy data is structured through a relational database or XML the task of data translation can be handled by a wrapper component.

In the case of relational databases some mapping tools like D2R or D2RQ can be use for direct translation into RDF format. If it is flat text, or mark-up such as HTML, (semi-)automatic content annotation can be made using an annotation manager component like

SMORE[2]. In these ways developed RDF data can be published using well-known web server like Apache or be using the specialized query interfaces like RDF NetAPI[3].

3.3.3 Requirements Analysis

In this case, our requirements are:

- Mapping tools to extract the RDF from structured data (not only relational databases but also XMLdocuments and other database types);
- Better annotation tools to (semi-) automatically annotate flat data using e.g. NLP techniques to identify relevant concepts and their properties.

3.4 Ontology Management

3.3.1 Problem Statement

The concept hierarchies used by the system, most significantly in the matchmaking task, are not expected to be static. Information in the job market is always subject to change, e.g. the classification of academic qualifications, classification of occupations or the requirements upon a particular profession (e.g. in computer science, new programming language, software or hardware specializations may be required). Hence the hierarchies must be updated while preserving the accuracy of the semantic matchmaking by the system.

3.3.2 Knowledge Processing Task and Component

The **ontology manager** component is responsible for providing the functionality of ontology updates and the system must be configurable to work with these changes.

3.3.3 Requirements Analysis

Best practises and guidelines are required for ontology changes, e.g. in this case how best to model new or changed concepts in the concept hierarchy to preserve accurate matching. This has to take into consideration the matchmaking approach in order to best determine the new positioning of concepts. Ontology changes should also be tracked to ensure a semantic alignment of information being generated by different sources, as it is likely over time that different information will be based on different versions of the original ontology.

3.4 Trust

3.3.1 Problem Statement

Not everything found in the Web is true and the Semantic Web does not change that in any way. From a trust perspective the participants form an open dynamic network of independent information providers. Hence all information found by the system must be considered a claim of its author rather than a fact. In the HR case the certificates and documents might be added to the applicant profile in the form of RDF-statements which would be signed by the appropriate organisation and institutions.

---

[2] Kalyanpur, A.; Parsia, B.; Hendler, J.; Golbeck, J.: SMORE – Semantic Markup, Ontology, and RDF Editor. http://www.mindswap.org/papers/SMORE.pdf, 2003,

[3] Moore, G.; Seaborne, A.: RDF Net API, W3C Member Submission. http://www.w3.org/Submission/2003/SUBM-rdf-netapi-20031002/, 2003

3.3.2 Knowledge Processing Task and Component

Trust could be considered in the tasks of results reconciliation, i.e. initial assessment of the trustworthiness of information before including it in the repository, of directory management, i.e. developing storage solutions which give precedence to trusted information, and of matching, i.e. using trust as a ranking criteria.

3.3.3 Requirements Analysis

A trust architecture is required which allows to decide which job applications or job postings are trustworthy. This may involve:

- determining the reputation of information providers.
- providing support for digital signatures, e.g. digitally signed testimonials.