Improving the recruitment process through ontology-based querying

Malgorzata Mochol¹, Holger Wache², Lyndon J. B. Nixon¹
holger@cs.vu.nl

¹Free University of Berlin
Networked Information Systems
http://ag-nbi.de

²Vrije Universiteit Amsterdam
Artificial Intelligence Section
http://www.cs.vu.nl/
Motivation
Project Context
e-Recruitment Nowadays
e-Recruitment Use Case
  - Requirements Analysis
  - Semantic Web-based Prototype
  - Extension of the Semantic Web-based Prototype
Conclusion & Future Work
Motivation

- Online Recruitment the main recruitment channel
  - 47% of German internet users (28% of European users) read the online job postings*
  - Over 50% of future employment procurement in Germany is expected to occur online

- Online personnel marketing = cost cutting and efficiency

- Maintenance of an overview of so many portals is a formidable task → visiting every job exchange site next to impossible
  - many websites and online portals financed by publishing fees
  - various business websites
  - portal set up by the state job centre

* “Daten für die Online-Rekrutierung I/2005” wwjGmbH
06.11.2006
Project Context

- **Knowledge Nets** (Wissensnetze)
  - Analysis of typical scenarios for the deployment of Semantic Web technologies
  - Prediction of the economic impact of Semantic Web technologies on e-Business
  - Combination of business and technology-driven analysis

- **European Network of Excellence Knowledge Web**
  - Goal is to achieve technology transfer from academia to industry
  - An Industry Area collects business use cases and promotes Semantic Web based solutions
  - Industry Portal [http://knowledgeweb.semanticweb.org/o2i](http://knowledgeweb.semanticweb.org/o2i)
Today’s e-Recruitment

- Job postings in the form of free text using uncontrolled vocabularies
- Meta-search engines – search on a full text basis
- Open positions published on employer websites → meta-job portals collect information from different sites
- General & specialized search engines as main tool in job search
  - 74% of internet users use various search engines as main tool for information retrieval*
  - 41.1% of internet users use Google*

- Problem: Search engines index job postings imprecisely:
  - Problem to recognize a job posting on the Web
  - Problem to extract relevant keywords (job title, skills, …) using linguistic methods
  - Search results limited in their ability to provide offers that match the precise needs

- Solution: Semantic annotation of job postings

*“Online-Rekrutierung I/2005” wwjGmbH
We are looking for a person which:

- has a degree in computer science
- wants to work in software consulting and development,
- is an expert in C, Java, PHP, UML, .Net and WindowsNT,
- has worked for at least 5 years in an industrial and 5 year in a research project,
- should have experience as project or team manager,
- should not be older then 25

"Online-Rekrutierung I/2005" wwjGmbH
06.11.2006
SEBIZ'06 - 1st International Workshop on Applications and Business Aspects of the Semantic Web
Semantic Web-based solution (I)

- **Human Resource Ontology** incorporates set of controlled shared vocabularies for employers, job applicants and job portals
Semantic Web-based solution (II)

- Semantic matching combines annotations using controlled vocabularies with background knowledge about the HR domain.
- Semantic search engines use the annotation to identify job postings.
- They could reliably crawl and index job postings.
- They offer semantic matching services to increase the precision of matching of open positions and applications.
Advantages of Semantic Web-based solution

● Employers could use semantic matching algorithms to automate the pre-selection of candidates.

● Job seekers could profit by increased transparency in the market.
### Semantic Web-based solution (III)

#### Details zu den Tätigkeiten (88.0% Ähnlichkeit, gewichtet mit 0.2)

**Wirtschaftszweig:** Softwareberatung und -entwicklung  
**Berufskennziffer:** Dipl.-Informatiker/in (Uni)  
**Bewerber:** Sonstige Softwareentwicklung  
**Ähnlichkeit:** 100.0%

**Bewerber:** Dipl.-Informatiker/in (FH) - Softwaretechnik  
**Ähnlichkeit:** 76.0%

#### Erforderliche Kompetenzen (90.7% Ähnlichkeit, gewichtet mit 0.6)

<table>
<thead>
<tr>
<th>Erforderliche Kompetenzen</th>
<th>Vorhandene Kompetenzen</th>
<th>Ähnlichkeit</th>
</tr>
</thead>
<tbody>
<tr>
<td>C# (Experte)</td>
<td>Java (Mittlere Kenntnisse)</td>
<td>89.7%</td>
</tr>
<tr>
<td>Java (Experte)</td>
<td>Java (Mittlere Kenntnisse)</td>
<td>92.0%</td>
</tr>
<tr>
<td>C++ (Experte)</td>
<td>Java (Mittlere Kenntnisse)</td>
<td>86.1%</td>
</tr>
<tr>
<td>Unix (Experte)</td>
<td>Linux (Experte)</td>
<td>92.1%</td>
</tr>
<tr>
<td>Servlet (Experte)</td>
<td>JSP (Experte)</td>
<td>92.1%</td>
</tr>
<tr>
<td>UML (Experte)</td>
<td>UML (Mittlere Kenntnisse)</td>
<td>92.0%</td>
</tr>
<tr>
<td>DotNET (Experte)</td>
<td>DCOM (Anfänger)</td>
<td>80.1%</td>
</tr>
<tr>
<td>JSP (Experte)</td>
<td>JSP (Experte)</td>
<td>100.0%</td>
</tr>
<tr>
<td>WindowsNT (Experte)</td>
<td>Linux (Experte)</td>
<td>92.1%</td>
</tr>
</tbody>
</table>
Some (still) open issues

- We are looking for a person which:
  - has a degree in computer science
  - wants to work in software consulting and development,
  - is an expert in .Net, JSP, UML, .Net, Unix and Windows,
  - has worked for at least 5 years in an industrial and 5 year in a research project,
  - should have experience as project or team manager,
  - should not be older than 25.
How to relax the query

- We are looking for a person which:
  - has a degree in computer science
  - wants to work in software consulting and development,
  - is an expert in C#, Java, C++, Servlet, JSP, UML, .Net, Unix and WindowsNT,
  - has worked for at least 5 years in an industrial and 5 years in a research project,
  - should have experience as project or team manager,
  - should not be older than 25

1. Another OO language would also be okay
2. Need not to have leading experiences but then he should be younger than 21
3. Drop the degree requirement
4. 3 years may also be enough
How to specify these relaxations

Skills
Experiences
Function
Age
How to specify these relaxations

**Rule 1**

If experience = JAVA
Then experience = OO-Language

**Rule 2**

If experience = X
Then experience = Y
With Y is father of X

IF experience = JAVA
THEN experience = OO-Language

Rule 1

IF experience = X
THEN experience = Y
WITH Y is father of X

Rule 2
Rewriting Rules (I)

When you find all patterns in a query ....

IF pattern\textsubscript{1}, pattern\textsubscript{2}, ...
THEN replacement\textsubscript{1}, replacement\textsubscript{2}, ...
WITH condition\textsubscript{1}, condition\textsubscript{2}, ...

... then substitute the patterns with these replacements ...

... but do the replacements only if the conditions are satisfied.

Very simple but expressive relaxation technique!
Replacing OWL queries

IF <owl:Class rdf:about="Java"/>
THEN <owl:Class rdf:about="PureObjectOrientedLanguages"/>
WITH true

IF <owl:Restriction>
  <owl:onProperty rdf:resource="#hasDuration"/>
  <owl:someValuesFrom>
    <owl:Class rdf:ID="FiveYearsOrMore"/>
  </owl:someValuesFrom>
</owl:Restriction>
THEN <owl:Restriction>
  <owl:onProperty rdf:resource="#hasDuration"/>
  <owl:someValuesFrom>
    <owl:Class rdf:ID="TwoYearsOrMore"/>
  </owl:someValuesFrom>
</owl:Restriction>
WITH true
Advantages and Disadvantages

**Sliders**

- "Estimates" the distance between my query and one result (provides a ranking)
- Sliders are difficult to use; relaxation steps have to be translated in an unnatural way

**Rewriting Rules**

- Easy to formulate the options how to relax a query.
- Returns a set of unordered results, i.e. the results can’t be ranked

**Combine both**
Proposal for Combination

Apply Slider technology in order to rank
• the clusters and
• the results in one cluster
Conclusion

- e-Recruitment is a natural application area for Semantic Web techniques
  - Job postings and offerings are annotated with metadata
  - Ability to search with complex and precise queries

- Over specified queries (normal case!) have to be relaxed
  - Sliders are inappropriate for representing relaxation steps but rank all results
  - Rewriting rules more human-oriented but can not rank the results
  - Combination of both most promising