

A Service of

ZBW

Leibniz-Informationszentrum Wirtschaft Leibniz Information Centre for Economics

Ionescu, Bogdan; Ionescu, Iuliana; Florescu, Vasile; Tinca, Andrei

Article

Semantic Annotation and Association of Web Documents: A Proposal for Semantic Modeling in the Context of E-Recruitment in the IT Field

Journal of Accounting and Management Information Systems (JAMIS)

Provided in Cooperation with:

The Bucharest University of Economic Studies

Suggested Citation: Ionescu, Bogdan; Ionescu, Iuliana; Florescu, Vasile; Tinca, Andrei (2012) : Semantic Annotation and Association of Web Documents: A Proposal for Semantic Modeling in the Context of E-Recruitment in the IT Field, Journal of Accounting and Management Information Systems (JAMIS), ISSN 2559-6004, Bucharest University of Economic Studies, Bucharest, Vol. 11, Iss. 1, pp. 76-96

This Version is available at: https://hdl.handle.net/10419/310481

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.



http://creativecommons.org/licenses/by/4.0/

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.



WWW.ECONSTOR.EU

SEMANTIC ANNOTATION AND ASSOCIATION OF WEB DOCUMENTS: A PROPOSAL FOR SEMANTIC MODELING IN THE CONTEXT OF E-RECRUITMENT IN THE IT FIELD

Bogdan IONESCU¹, Iuliana IONESCU, Vasile FLORESCU and Andrei TINCA The Bucharest Academy of Economic Studies, Romania

ABSTRACT

The emergence of the Internet has widely opened opportunities for web applications based on the modeling of textual knowledge from electronic documents, with the aim of identifying the useful parts. This is also the case of e-recruitment applications, which aim at the optimizing and automating the processes by semantic association between CVs a job offers. Our paper considers the problem of semantic-based annotation of semi-structured documents, and proposes a semantic model for e-recruitment in the IT domain.

e-recruitment, ontology, semantic annotation, information system, similarity measure

INTRODUCTION

The Internet has become essential, and the web is currently the dominant paradigm for the optimization of the recruitment processes. Most of the job seekers post CVs on web servers and employers post job offers. The employment market is gradually moving onto the web using semi-structured documents (Rafter & Smyth, 2000: 1; Bizer *et al.*, 2005: 1; Kessler *et al.*, 2009: 2; Kessler, 2010: 16; Popescu & Popescu, 2010: 1), which translates into massive databases containing CVs and job offers, which are difficult to process in the absence of adequate techniques. This is the reason why research is done in the area of optimizing recruitment processes using Internet technologies, by adding semantics to standardized documents containing information regarding CVs and job offers (*Figure 1*).

Correspondence address: Bogdan Ionescu, The Faculty of Accounting and Management Information Systems, ASE Bucharest, 6 Piata Romana, Bucharest, Romania, email: ionescub@gmail.com

The semi-structured documents are semantically annotated (based on their content) and they are related to a field ontology. In the particular case of recruitment, the field ontology is inspired from the most significant parts from CVs and job offers (Yahiaoui *et al.*, 2006). "Intelligent" programs can compare document instances using the common reference (ontology) and also detect ontological similarities between multiple instances.

Despite significant progress in the area of semantic annotation of semi-structured documents, the process of creating the annotations and particularly the matching of these documents are still difficult and resource-intensive. An essential factor to the recruitment process is the creation of a *controlled vocabulary* for the field, and the construction, maintenance and update of the ontology (concepts, properties and relations). These two elements guide the process of data extraction from semi-structured documents and thus ease the process of their semantic annotation (as compared to annotations generated in the absence of ontology).



Figure 1. Generic framework for the recruitment process

Our paper aims to address the problem of semantic-based annotation of semistructured documents, and proposes a semantic model for e-recruitment in the IT domain.

The paper is structured as follows: Section 2 presents the research methodology. Section 3 presents the literature review in the area of semantic annotation of semistructured documents related to the recruitment process (CVs and job offers) and their pairing using an ontology. Section 4 proposes a global framework for modeling the ontology of e-recruitment in the IT field. We conclude with the last section, where we consider the implementation of the proposed models by developing an e-recruitment platform.

1. RESEARCH METHODOLOGY

In our research, we have taken an action-based, constructivist approach: the literature review (where the fundamentals of semantic-based modeling for e-recruitment documents) is followed by the proposal of a controlled vocabulary and an ontology for the semantic annotation of recruitment documents in the IT field.

The semantic-based approach to the optimization of data-flows within the human resources (HR) domain, in connection with the profiles of the job offers, is followed by our conceptual clarifications regarding the process of annotation and pairing of semi-structured documents. The results of our research recommend a new approach to human resources management, and offer support for developing an e-recruitment platform, founded on the utilization of the field ontology in the processes of annotation and semantic pairing of semi-structured documents.

2. LITERATURE REVIEW

2.1. E-recruitment and the job market

The recruitment process is based on the sum of the actions required to find candidates matching the job offers. E-recruitment involves the use of IT and communications technologies, and particularly the web (Bizer et al., 2005: 2, Trichet & Radevski, 2006: 1; Yahiaoui et al., 2006: 2). However, e-recruitment does not yet have a strictly-defined sense and understanding (Fondeur, 2006: 2; Mellet, 2006: 3). As a "cousin" of e-business, it refers to the instruments used to facilitate and eventually automate the web-based recruitment process. According to Kessler et al. (2008: 1), there was a quick transition from pure intermediary, informative processes to externalizing the recruitment towards entities specialized in e-recruitment. The success enjoyed by e-recruitment has led to the creation of many job portals, also called job boards (Yahiaoui et al., 2006: 2; Kessler et al, 2009: 3; Popescu & Popescu, 2010:2). This situation created awareness and risen the interest in the modeling and semantic search of semi-structured web documents.

The formal modeling of a document's contents in terms of *obtained* (as is the case with CVs) and *required* (job offers) with the help of an ontology, between profiles for applicants and work requirements, opens the possibility to automate the processes of comparing, matching and associating these documents on a semantic basis, yielding a matching coefficient (Trichet & Radevski, 2006: 2; Kessler *et al.*, 2008: 3; Kessler *et al.*, 2009: 3; Thiam, 2010: 5).

Vol. 11, No. 1

2.2. The Annotation and Association of Web documents, based on content

2.2.1. General Framework

The semantic annotation of web documents (*Figure 2*) is essential for their storage, as well as the search and retrieval of the documents relevant to a particular need (Hernadez & Aussenac-Gilles, 2004: 3; Abrouk, 2006: 5; Thiam, 2010: 5).

By semantic annotation, we attach a *"note"* to semi-structured documents; in other words, meta-data which describes the contents both in a formally and explicitly. This note can be stored inside the document or in another document related to the content through an URI (Uniform Resource Identifier), being available to users or agent programs with the purpose of finding documents relevant for a certain requirement (Abrouk, 2006:7). The construction of the note is based on a thesaurus, or a field ontology (Guissé *et al.,* 2009:4), both founded on a controlled vocabulary. A variety of programs which annotate semantically are already available (Aussenac-Gilles *et al.,* 2008: 3; Thiam, 2010: 6). Mathet and Widlöcher (2009: 2-3) introduce generic annotation instruments such as "Glozz", a platform which processes linguistic objects (especially discursive).

Figure 2. Framework for semantic annotation and pairing



A controlled vocabulary is defined as a collection of terms, defined by experts in a field (Hernandez, 2005: 4). The term is a word (e.g., State budget), or a group of words (e.g., Public debt). The meaning of the words is not necessarily defined, and

there is no logical connection between the terms. There are two broad categories of terms:

- Descriptors, which explicitly represent a concept contained in a reference text;
- Non-descriptors (synonyms), always linked to a descriptor

The controlled vocabulary can be used to label content. A catalogue is a typical example of a controlled vocabulary. A *thesaurus* is a hierarchical dictionary of normalized terms, also called descriptors. A descriptor shows the context for term usage, and has a unique, non-ambiguous sense (multiple senses and synonyms are eliminated).

- A *lexicon* contains descriptors (representing a concept of the field) and nondescriptors (linked to a descriptor, generally synonyms);
- A collection of *definitions*, also known as *notes*;
- A classifying structure (also known as a *direct semantic medium*), expressed as a structure of relations between a lexicon's terms.

In Figure 3 we present the classifying structure for Human Resources management, according to the on-line thesaurus MOTBIS (2011). For the description of the semantic medium we use several types of descriptors: generic (GD), specific (SD), equivalent (ED), and associative (AD).





(Adapted from MOTBIS, 2010:on-line rendering)

Vol. 11, No. 1

A *field ontology* corresponds to a generic, conceptual description of the entities belonging to a field, and is required for building applications based on knowledge. An ontology is composed of classes and entities. The classes are composed of entities which are similar. A machine-understandable class id equivalent to the "term" concept, as understood by a human. Thus, a field ontology consists of concepts specific to a particular domain. The field ontology is superior to a thesaurus from a conceptual and semantic point of view, but some of the hierarchies stem from the thesaurus. In practice, an ontology is a graph structure, where the nodes represent the concepts, and the links represent the relations (or roles) between concepts. The generic framework for semantic annotation is presented in Figure 4.





(Source: adapted from Yahiaoui & Boufaida, 2006: 8)

A field ontology can be encoded using OWL (Web Ontology Language), which is based on XML. Several editing programs are available (TERMINAE, Differential Outology Editor, Ontology Editor, Protege2000, Ontolingua, etc.).

The semantic structures (meta-data) obtained through this guided extraction (using a thesaurus or ontolgy) are associated to semi-structured documents (Desmontils & Jacquin, 2002: 3). The documents are retrieved according to the link between the semantic structure and the source document.

The semantic indexing of semi-structured documents is based on the meta-data extracted from the sources. The indexing facilitates semantic queries issued for a specific order. According to Guissé *et al.* (2009: 2) the indexing is used to describe the sum of the semantic annotation, seen as a space between texts and a semantic structure. The authors differentiate three sub-structures for an index, which describe:

• Which text fragments are documentary units (*du*), which can be used to establish an indexing link;

- Which semantic units (*su*) can be associated to documentary units, and which relations these units have (a thesaurus or an ontology);
- The association of documentary units to the semantic units(du_i , su_i), in other words, an indexing link.

Generally, an indexing link can be represented as $(l_i; du_i; su_i)$, where l_i is the list of properties linking du_i to su_i . The structure of the index is in accordance with the structure of the ontology for the field where the documents belong. The semantic indexing of semi-structured web documents can be realized:

- Automatically, by determining in an automated way the most important terms of the document, for further analysis. This is how search engines process Web pages, by "crawling" the web and associating an index to pages;
- Semi-automatically, using specific techniques for knowledge engineering and automated treatment of language documents (Desmontils & Jacquin, 2002: 2), having a human who aides and guides the process.

The similarity between the concepts of an ontology, calculated from the taxonomical link « is-a », can form the basis of XML document indexing (Zargayouna & Salotti, 2004: 5).

The *semantic querying* of databases containing semi-structured XML documents is based on using the ontologies and identifying the similarities (Bizer *et al.*, 2005: 7; Boucetta *et al.*, 2005:4) in: ontologies, queries and semi-structured XML documents, and in semi-structured XML documents.

2.2.2. Semantic annotation of CVs and Job Offers

The first step in the annotation process is the creation of a HR ontology. Currently, the most widely used standard for publishing job offers and applications is HR-XML, which is developed and maintained by the HR Consortium (which is an independent association dedicated to improving the automation of human-resources related data-exchanges through the use of XML.) Because the hiring process is no longer restricted to a national level, the process of semantic indexing must rely on the occupational classifications specific to a certain language/geographical zone. In the US, according to the Standard Occupational Classification, the workers are classified into 840 occupations (23 major groups and 97 minor groups). In the Euro area, given the differences in languages and education systems, there are ongoing efforts to level the playing field. The process is set up so that each country must have its input into the European Qualifications Framework. This framework will provide a common reference used to asses knowledge, skills and competences.

Vol. 11, No. 1

According to Bizer *et al.* (2005: 9), the HR ontology is split into several subontologies, as follows:



Figure 5. Sub-ontologies within the HR Ontology

Each sub-ontology must be modeled taking into account the particularities of the national standards, and we must mention that the development of such standards requires considerable effort and expert skills. Also, the ontologies must be permanently updated, which is a difficult process when multiple national actors are involved, such as in the EU.

2.2.3. Semantic Matching between CVs and Job Offers

Semantic matching is the process in which the metadata extracted from a CV is compared to the metadata of a job offer, using a controlled vocabulary associated to the domain. The result of the matching process must be a Matching Index (MI).

The MI indicates the quality of the match, by comparing several concept hierarchies (such as qualifications and competencies) expressed as a tree structure (Bizer *et al.*, 2005: 10).

Besides the basic requirements, there are several criteria which should be satisfied by the matching process:

- It must account for employer preferences regarding the weighting of competencies expressed in the job offer;
- It should verify the credentials presented in a CV, based on evidence such a electronically signed documents. The quality of the credentials should influence the weighting.

⁽*Source:* Bizer *et al.*, 2005: 9)



Figure 6. The process of annotation and matching

The following example relies on ideas from Bizer *et. al.*, (2005: 11), and Boucetta *et. al.* (2005: 9). To measure the similarity between two concepts we must calculate the distance (*d*) between the concepts, as represented by their position in the hierarchy. For concepts c_1 and c_2 , their similarity is expressed as $sim_c(c_1, c_2)=1-d_c(c_1, c_2)$. However, the distance between concepts situated higher in the hierarchy (for example, *medicine* and *IT*) is greater than the distance between concepts lower in the hierarchy (*databases* and *programming*). To account for this, a *milestone* is attributed to each level, as follows:

$$milestone(n) = \frac{1/2}{k^{l(n)}}$$

k is a factor larger than 1 which indicates the rate at which the value decreases throughout the hierarchy, and $l_{(n)}$ is the "depth" of the node in the hierarchy, as illustrated in Figure 7.

In our example, we have set k = 2, and for the root, we consider $l_{(n)} = 0$. We note that the path between any two concepts on the graph passes through a common parent, and so, the distance between concepts will be measured by their milestones and the milestone of their closest parent:

$$\begin{cases} d_{c}(c_{1}, c_{2})^{-} d_{c}(c_{1}, ccp) + d_{c}(c_{2}, ccp) \\ d_{c}(c, ccp) = milestone(ccp) - milestone(cc) \end{cases}$$

where ccp is the closest common parent of c_1 and c_2 . This model supports the assumption that the distance between close "brothers" (like "C++" and "Java") is smaller than the distance between "cousins" (such as "Invoicing" and "Oracle Databases").

Vol. 11, No. 1



Figure 7. **Ontology component with associated milestone values**

(Source: adapted from Bizer et. al., 2005: 11)

To illustrate, we wish to calculate the distance between the concepts of "MySQL" and "C++". The closest common parent is "IT", and thus the calculation is:

 $d_{c}(MySQL, C++) = d_{c}(MySQL, IT) + d_{c}(C++, IT)$ = (1/4-1/16) + (1/4-1/16) = 0,25

Thus, the similarity between the two concepts is:

 $sim_{c}(MySQL, C++) = 1 - 0.25 = 0.75$

The distance between two remote concepts would be:

 $d_{\rm c}$ (Invoicing, OracleDatabase) = (1/2-1/16) + (1/2-1/16) = 0,875

and their similarity:

 sim_c (Invoicing, OracleDatabase) = 1-0,875 = 0,125

Bizer et. (2005:11) al propose a scheme for integrating competence levels *(cl)* required by a job offer using the following formula:

$$sim_{p}(cl_{1},cl_{2}) = \begin{cases} 1-\alpha(cl_{1}-cl_{2}) & |cl_{1}-cl_{2}\geq 0\\ 1 & |cl_{1}-cl_{2}< 0 \end{cases}$$

Where $0 < \alpha \le 0.25$ is a factor which indicates the rate at which the similarity sim_p decreases with the increasing gap in competence levels. For example, assume an expert in databases has cl=5 and a novice in programming has cl=1. Setting $\alpha=0.2$, we have $sim_p = 1-0.25(5-1) = 0.2$. This value, multiplied with the concept similarity (calculated at 0.25) yields sim_c (databases, programming)* sim_p (expert,novice)=0.2*0.25=0.05. This insures that similarity between more important concepts has more weight in the final calculation.

Therefore, the formula for calculating the matching index (MI) between a job offer (jo) and a job seeker (cv) is:

$$MI(jo,cv) = \sum_{i \in I} w(c_{jo}^{i}) * \max\{sim_{c}(c_{jo}^{i}, c_{cv}^{i}) * sim_{p}(p(c_{jo}^{i}), p(c_{cv}^{i})| j \in J\}$$

where $\sum_{i \in I} w(c_{ig}^{i}) = 1$

Each skill from the job offer (c_{jc}^{t}) is compared with each skill in the cv (c_{cv}^{t}) , taking in account both concept and required competence similarity. The best match is multiplied by the weight factor; the sum represents the final matching index.

2.2.4. Support system for semantic annotation and matching of CVs and job offers

In the current context of an increasingly virtual job market, a multitude of instruments support systems based on ontologies are developed, with the aim of enhancing the data-flows. In Figure 8 we present the general architecture of an annotation and matching (pairing) system for e-recruitment, relying on the human resources ontology $(HR_Ontology)$.





(Source: adaptation from Yahiaoui et al., 2006: 3 and Boucetta et al., 2008: 3)

Vol. 11, No. 1

The actors of such a system (job seekers, employers, recruiters) are assisted in erecruitment processes for the following types of activities:

- Semi-structured documents are uploaded on the web (CVs, job offers) in a XML data repository
- The documents undergo the process of semantic annotation using the HR ontology, the result of this process is meta-data which are stored together with the documents
- A user launches a query to determine the matches found on the server, the query is formalized and forwarded to the association component, which relies on meta-data and on the ontology, using an inference engine. The result expresses the association (match) index, and the elements specific to the document (URI, matching index according to: personal qualifications, requirements, and competencies). The result is presented as a "quadrulpe", assisting the user in making a decision.

The human resource ontology is constructed for the semantic annotation of CVs and job offers. Its concepts and relations between concepts are inspired from the common parts of the most significant sections of CVs and job offers (personal qualifications, requirements, diplomas, personal experience and competency).

3. PROPOSAL FOR AN E-RECRUITMENT ONTOLOGY, WITH AN EXAMPLE OF AN ONTOLOGY IN THE DATABASE AREA

3.1. Modeling the common reference for e-recruitment in the IT area

The IT field is of special interest to large corporations, from a human resource point of view. The optimization of the e-recruitment processes involves, firstly, a common reference (ontology) able to link the employer requirements and job seekers' offers, namely, what the job seekers obtained by studies and professional experience. In our research we opted for three types of ontologies on which the recruitment process should be based (*Figure 9*):

- Job seekers based data extracted from CVs
- Employers based on requirements related to the job offers
- The national accreditation of qualifications and occupational standards based on a national occupational classification

The competencies represent the shared part for the three types of ontologies, and at the same time the fundamental component for the development of a support system in the e-recruitment processes.



Figure 9. Generic framework for the definition of IT ontologies

3.2. IT competency ontologies within job offers

For the IT field, regarding the job seekers, we define a generic ontology named Competencies. Within this competency, we find the IT Competency sub-ontology, present in the majority of the CVs and job requirements issued by employers. This sub-ontology is composed from knowledge and technical skills found in the IT field, and implemented in the corporate IT systems.

According to the specific requirements found in job offers, some IT Competency subontologies can be considered as complementary to main (specific) ontologies, related to well-defined activity areas (accounting, analysis, marketing etc.), in connection to the specificity of the job offers. However, for the IT field, the sub-ontologies stemming from "IT Competencies" can be considered of special importance.

For the field of IT, three sub-ontologies are of interest, and each of them can be further organized into sub-ontologies:

• Theory and Technical Foundations - these define IT specific competencies, both from a theoretical and practical point of view. Some examples are: technical and functional architecture of IT systems, the design and development of IT systems, database management, IT governance, data and application security, communication, networking etc.

- **Organizational competencies** these define the elements of organizational culture, adjacent to the IT field; these are required for the implementation, integration and maintenance of IT systems in the corporate environment. Some of these competencies gaining traction in the IT field are: audit knowledge, accounting, activity budgeting, human resource and occupational standards management, quality management, legal knowledge, etc.
- **Technical skills** referring to one's capacity to use technologies and products, and behavioral skills, referring to action and anticipation, pedagogy, problem-solving, professional efficiency (adaptability, integrity, pragmatism, rigor), management skills (leadership, management, organization), and relational (good listener, communication), etc.

3.3. Ontologies based on employer requirements within job offers

The employers specify, for job openings in the IT area, requirements focused mainly on competencies, in correlation with national and European occupational standards. In the IT domain, the job offers fall within the following sections:

- IT systems management
- IT project management
- The design, development, implementation and maintenance of an IT system;
- The study, design, implementation, integration and operation of IT infrastructure
- Technical support offered to users
- Control, audit, and IT security
- Operational management—for IT employers (budget, decision)

Each sub-ontology can be developed on other sub-ontologies (or subjects), which, at this level, are essentially job offers (which must be issued in accordance with the national occupational standard). In turn, these offers are associated to sub-ontologies described in the competency sections relating to applicants' CVs.

For example, for the sub-ontology "The design, development, implementation and maintenance of an IT system", which actually represents the lifecycle of software products, we can define the next positions:

- Specialist for currently-implemented IT systems;
- IT systems designer;
- IT systems tester;
- Application integrator;
- Product customization

3.4. Ontologies based on occupational standards

The ontologies based on occupational standards are issued by national entities which define and implement occupational standards. Periodically, these entities publish a classification which contains professional competencies required by employers to recruit qualified work force.

This classification is the result of the experience of human resources directors from the IT field, and aims to be a formal instrument for a common description of employment requirements in professions related to organizations within the field of informational systems. Essentially, this occupational standard classification offers a clear image of the evolution of corporate IT systems. The formalization of the professions and occupational standards related to corporate IT systems is a procedure which is frequently updated due to the constantly changing nature of the IT field, which brings along structural modifications in the requirements for IT jobs.

The common reference for occupational standards in the IT recruitment field has undergone several mutations recently, due to the orientation towards "IT competencies", complementary to occupational standards based on jobs. This new orientation brings the possibility if an international reference for IT competencies (e-Competence Framework) which, is implemented in Europe by CNN/ISSS (European Committee for Standardisation / Information Society Standardisation System).

The structuring of competencies and occupational standards in the IT field is, in essence, a technical view of an information system. The occupational standards specific to IT will undergo structural changes, moving away from an emphasis on technical aspects related to the IT jobs, towards a global view, in which corporate IT systems are directly linked to the business and its strategy, and that IT systems are not only technical, but their role is strongly connected with enterprise functions.

In Romania, eighteen occupational standards are defined in the field of IT and telecoms (CNFPA, 2010):

- 1. Application administrator
- 2. Computer and network operator
- 3. Data input technician
- 4. Text and image manipulation technician
- 5. Operator in the field of computer-assisted design
- 6. Specialist in the field of computer-assisted design
- 7. Graphic designer (DTP designer);
- 8. Web designer (high-school level studies);
- 9. Systems software engineer
- 10. Computer network administrator
- 11. Information systems programmer
- 12. Auxiliary programmer
- 13. Information system consultant

Vol. 11, No. 1

- 14. IT department directors
- 15. Analyst
- 16. Information systems architect
- 17. Database administrator
- 18. Specialist in IT security and processes

These occupational standards, which in light of the semantic web are actually ontologies, are not currently associated to domains specific to IT.

3.5. Formalizing an unitary ontology for e-recruitment in the IT field

For the design of an e-recruitment portal, all the components based on competencies—offered by the job seekers and required by employers, and included in the Romanian occupational standards, must be included in an ontology—even if the components are not hierarchical. Thus, in the current paper, we propose to develop the sub-ontology related to IT scientific/technical competencies.

This IT scientific/technical competencies ontology can be hierarchically organized in more sub-ontologies which actually define information technologies, useful both for research and in the corporate environment. In building such a hierarchy we must cover all IT domains according to the related disciplines, from an applied science point of view. The correctness of such an undertaking depends on the association between the know-how of the informational field and the ontologies found in CVs and required by employers in the recruitment process.

We briefly introduce some semantic aggregations related to the IT field:

- E-learning;
- Intelligent systems;
- Databases;
- Management information systems;
- IT&C;
- Web Based Communities;
- Interfaces and Human Computer Interaction;
- Data Mining;
- Networks; e-technologies;
- Computer Graphics;
- Collaborative Technologies;
- Informatics, etc.

Each sub-ontology can be placed in a hierarchy according to the "Theory Foundations" and "Technical aspects" criteria. We present some examples of sub-sub-ontologies, according to the aforementioned criteria, for the field *Informatics*:

1) **Theory Foundations**: algorithms; architectures; Artificial intelligence; Compilers; Complex Systems; Data modeling; Expert Systems; Interfaces; Numerical computation; Object Orientation; Ontologies; Programming

Languages; Programming Techniques; Scientific Computing; Service Oriented Architecture (SOA)

2) **Technical aspects**: Adaptive Systems; Computer Aided Design; Computing Practices; Database Management; Embedded Systems; Interoperability; Networking; Simulation; Software Development; Software Engineering; System Integration; UML.

The sub-ontologies are quantified by attributed which are finally used to form scores, by successive aggregations. These are generally attributed by the employer, according to the importance defined by the requirements of the job offer. At an atomic level, the ontologies contain themes, or subjects. These themes describe the last level of sub-ontology in the hierarchy.

3.6 Sub-ontologies applied to the database field

For the IT field we propose a number of significant semantic aggregations of a subontology which describes the necessary competencies for a recruitment candidate, required by the employer for a job in the field of databases (*Figure 10*). According to this target, the database sub-ontology was divided in more sub-ontologies, according to "Theory Foundations" and "Technical aspects". Thus, two new sub-ontologies emerged, namely: "Data Modeling", quantified with a score of 30%, and "Database Management", quantified with 45%.



Figure 10. Ontology for the database field

While the majority of the sub-ontologies belong almost exclusively to the Database sub-ontology, more precisely, to "Theory Foundations" and "Technical aspects", some of them belong to the fundamental domain of database utilization ("XML and databases" and "Designing a Business Intelligence Solutions").

In parallel with the definition of sub-ontologies directly referring to fundamental and technical aspects related to databases, according to employer requirements, more items were required, such as "Working with operating systems", "Programming concepts", "Understanding of server management". These sub-ontologies belong to other ontologies, related to distinct areas within the IT field, such as: operating systems architecture, programming languages and physical server architecture.

We present an example of ontologies for the database domain (100%):

- Working with operating systems (5%)
- Programming concepts (10%)
- Understanding of server management concepts such as roles and tracing (10%)
- XML and databases (5%)
 - Describing XML functionality
 - Shredding XML data
 - Working with XML data types
 - Applying data integrity to XML data
 - Describing best practices for working with XML data
- DATA Modeling (30%)
 - > Core relational database concepts (10%)
 - > Designing databases at both the conceptual and logical levels (10%)
 - Data Integrity Using Constraints (5%)
 - Describing data integrity
 - Enforcing domain integrity
 - Enforcing entity integrity
 - Enforcing referential integrity
 - DATABASE MANAGEMENT (45%)
 - > Implementing DataBases (15%)
 - Work with the data types
 - Design and implement tables and views
 - Manage the concept of an index.
 - Design and implement stored procedures.
 - Managing and Monitoring Transactions
 - Implement table types, table valued parameters.
 - Describe transactions.
 - Querying Relational Data (15%)
 - Retrieve data.
 - Explain and perform single table queries.

- Perform joins between tables.
- Summarize data and aggregate functions.
- Modify data.
- Write a sub-query.
- Use Stored Procedures.
- Database administration (10%)
 - Administer server and surface security, access, and network configuration; Import, export, transform, and replicate data
 - Manipulate schemas, tables, indexes, and views
 - Automate maintenance and implement policy-based management
 - Monitor server activity and tune performance
 - Manage log shipping and database mirroring
 - Perform backups and recovery.

Designing a Business Intelligence Solutions (5%)

- Design a Business Intelligence (BI) architecture in client-server databases.
- Design a strategy for implementing the extract, transform, load (ETL).
- Design a strategy for managing packages.
- Design an online analytical processing (OLAP) solution architecture.
- Design queries for an OLAP solution.
- Design and develop a Reporting solution architecture.
- Design data mining solutions.

CONCLUSIONS

The topic of semi-structured web document annotation, with the aim of finding and associating the relevant documents, is of interest to academia, but also to the job market. Our paper starts with a literature review, then focuses on the semantic modeling of e-recruitment documents, then proposes an ontology for the database field, and concludes with further research topics. The proposed ontology is inspired from the common parts, which we considered significant, of CVs and job offers in the field of databases.

The paper also brings clarifications regarding the essential concepts for the semantic modeling of e-recruitment systems: field ontology, semantic annotation, semantic indexing, and semantic association of documents. For the continuity of the research, the authors envision three objectives. The first objective is to design an e-recruitment system, integrating the proposed models. The second objective is the development of a support system for semantic annotation of semi-structured documents related to e-recruitment processes. The third objective is to add new functionality to e-recruitment systems, increasing the relevance of meta-data extracted from the CVs and job offers, by offering advice to job seekers based on their profiles, and helping employers formulate better job offers.

REFERENCES

- Abrouk., L. (2006) "Annotation de documents par le contexte de citation basée sur une ontologie", Thesis, Universite Montpelier II, available on-line at http://tel.archives-ouvertes.fr/docs/00/14/25/68/PDF/these.pdf
- Aussenac-Gilles, N., Despres, S. & Szulman, S. (2008) "The terminae method and platform for ontology engineering from texts", in P. Buitelaar & P. Cimiano (Eds.), Bridging the Gap between Text and Knowledge: Selected Contributions to Ontology learning from Text, IOS Press
- Bizer, C., Heese, R., Mochol, M., Oldakowski, R., Tolksdorf, R. & Eckstein, R. (2005) "The Impact of Semantic Web Technologies on Job Recruitment Processes", WIRTSCHAFTSINFORMATIK, Part 15: 1367-1381
- Boucetta, Z., Boufaida, Z. & Yahiaoui, L. (2008) "Appariement sémantique des documents à base d'ontologie pour le E-recrutement", available on-line at http://www.google.com/url?sa=D&q=http://www.google.com/search%3Fq%3D cache:http://liris.cnrs.fr/Documents/Liris-4967.pdf&usg=AFQjCNEAYLZs 5G9lcV sFckLxS5ZcH5P9Q
- Desmontils, E., Jacquin, C. & Morin, E. (2002) "Indexation sémantique de documents sur le Web: application aux ressources humaines", available on-line at http://enssibal.enssib.fr/autressites/RTP/websemantique/octobre/octobre2/ Jacquin.pdf
- Fondeur, Y. (2006) "Internet, recrutement et recherche d'emploi : une introduction", Revue de l'IRES - numéro spécial : Internet, recrutement et recherche d'emploi, vol. 3, no. 52: 3–10
- Guissé, A., Lévy, F., Nazarenko, A. & Szulman, S. (2009) "Annotation sémantique pour l'indexation de règles métiers", TIA 2009, available on-line at http://www.irit.fr/TIA09/thekey/articles/guisse-levy-nazarenko-szulman.pdf
- Hernandez, N. (2005) "Ontologies de domaine pour la modélisation du contexte en recherche d'information", *Doctoral Thesis, Université de Toulouse*
- Hernandez, N. & Aussenac-Giles, N. (2004) "OntoExplo : Ontologies pour l'aide à une activité de veille ou d'exploration d'un domaine", *Actes de la VIème Journées de l'innovation*
- Kessler R., Bechet N., Roche M., El-Beze M. & Torres-Moreno, J.M. (2008) "Automatic profiling system for ranking candidates answers in human resources", OTM '08, Monterrey, Mexico, p. 625–634.
- Kessler, R. (2010) "Traitement automatique d'informations appliqué aux ressources humaines", Thesis , Academie D'Aix Marseilles
- Kessler, R., Béchet, N., Torres-Moreno, J.M., Roche, M. & El-Bèze M. (2009) "Profilage de candidatures assisté par Relevance Feedback", TALN 2009, Senlis, 24–26 juin 2009, available on-line at http://academic.research. microsoft.com/Publication/11024713/profilage-de-candidaturesassist%C3%A9-par-relevance-feedback

- Mathet, Y. & Widlöcher, A. (2009) "La plate-forme d'annotation Glozz: environnement d'annotation et d'exploration de corpus", *Actes de TALN 2009, Senlis, France*
- Mellet, K. (2006) "Sésame, ouvre-toi ! Analyse des données d'usage d'un moteur de recherche d'annonces d'offres d'emploi : www.keljob.com", *Revue de l'IRES-numéro spécial: Internet, recrutement et recherche d'emploi*, vol. 3, no. 52: 71–100
- MOTBIS (2011) available on-line at: http://www.motbis.cndp.fr/index.php/indexergeneralites/24-regles-dindexation.html
- Popescu, M. & Popescu, E. (2010) "A Human Resource Ontology for Recruitment", Faculty of Economic Sciences, available on-line at http://www.fse.tibiscus.ro/ anale/Lucrari2010/155.%20Craioveanu%20Georgiana.pdf
- Rafter, R. & Smyth, B. (2000) "Passive Profiling from Server Logs in an Online Recruitment Environment", Proceedings of the IJCAI Workshop on Intelligent Techniques for Web Personalisation 2001, available on-line at http://maya.cs.depaul.edu/~mobasher/itwp01/papers/rafter.pdf
- Thiam, M. (2010) "Annotation Semantique de Documents Semi-structures pour la Recherche d'Information", These, available on-line at http://tel.archivesouvertes.fr/docs/00/54/29/32/PDF/These MT8dec2010.pdf
- Trichet, F. & Radevski, V. (2006) "Competency-based Systems dedicated to e-Recruitment", Proceedings of the 4th International Conference on Computer Science and Information Technology, Amman (CSIT'2006), vol. 3: 185-197
- Yahiaoui, L., Boufaïda, Z. & Prié, Y. (2006) "Automatisation du e-recrutement dans le cadre du web sémantique", Proceeding of the 17th Journées Francophones d'Ingénierie des Connaissances (IC 2006), 28-30 June 2006, France
- Yahiaoui, L. & Boufaïda, Z. (2006) "Annotation semantique de documents Application au e-recrutement", available on-line at http://bu.umc.edu.dz/ theses/informatique/YAH4491.pdf
- Zargayouna, H. & Salotti, S. (2004) "Mesure de similarité dans une ontologie pour l'indexation sémantique de documents XML", available on-line at http://hal.archives-ouvertes.fr/docs/00/38/05/73/PDF/IC2004_Zargayouna.pdf

Vol. 11, No. 1