Semantic Characterization of Academic and Occupational Profiles Based on Competencies

Alexandra González Eras[1] Jose Aguilar [1,2]
[1] Computer Sciences Department, Universidad Tecnica Particular de Loja, Ecuador
[2] CEMISID, Engineering Faculty, Universidad de Los Andes, Mérida, Venezuela
acgonzalez@utp.edu.ec, aguilar@ula.ve

***Abstract** - This research shows an architecture for extraction, comparison and feedback of competencies from the characterization of its components. The main product is a scheme that facilitates skills and knowledge detection in documents, as well as the identification of more complex concepts, such as competencies. This research is a first contribution to the development of a system for comparison and update profiles, that is adaptable to the context, and facilitates understanding of the competency dynamics on education and employment environments.*

Keywords: Natural Language Processing, linguistic patterns, Competency.

1 Introduction

For universities and employers manage the competencies is a two way process, including skills identification to create new graduate profiles and determine qualified professional skills to fill a working place. However, in reality it is almost impossible to compare competencies, mainly due to incompatible profiles [Fazel-Zarandi and Fox, 2009] [Stevens, 2013]. In addition, the information published on university websites and work platforms is unstructured, ambiguous, and sometimes incomplete [Fazel-Zarandi, 2013]. Looking for a solution, models and platforms have been proposed in order to profiles standardization [Draganidis and Mentzas, 2006] and comparison through competency frameworks[1]. Nevertheless, these tools are rarely used by the actors, or it have only been proposed for one language context without a real application in others. As a result, universities can not identify the requirements of employers, and in turn, employers can not identify new graduate profiles aligned with their job offers.

The goal is to establish a management system of academic and labor profiles, allowing extraction, comparison and update of competencies, based on components characterization. First, we give an overview of the system architecture, and then focus on the characterization module, which models skills and knowledge, and proposes how they are connected to form more complex concepts, such as competencies. The resulting scheme is the management system cornerstone, which is applied on competency profiles in Spanish.

Notably, we initially focused on modeling skills and knowledge, because they are the most common concepts associated with the definition of competency in both contexts: labor and academic; in addition, skill and knowledge patterns are trackable in the profiles. This allows us to perform a less subjective characterization of competencies compared to other constituent elements, such as actitude, and value [Yahiaoui et al., 2006].

Thus, this research displays a overview of the system architecture detailing the characterization module, then we conducted two experiments, one on competency profiles to determine the ability of the scheme in recognizing patterns, and second analyses the ability of patterns to comparing the profiles against two bodies of knowledge (SWEBOK and DISCO). Finally, we made preliminary conclusions and future work of our research.

2 Academia versus occupational environment

Training for work is a mix between education, work experience and specific training acquired throughout life, hence competencies are defined and constructed in social practice, as a joint effort between companies, workers and educators. From a business standpoint, c*ompetence management* is the systematic development of human resources in organizations, and in that sense, competence management systems should support this systematic development [Lindgren et al., 2004].

Typically, professional skills and knowledge are adquiridad at university, where students develop a competence profile. In [Dorn and Pichlmair, 2007] states that through this profile, we may perform a gap analysis of the student, and investigate which additional knowledge and skills should be acquired, to achieve professional profile defined by the university or industry groups. Also, the student can use the profile to create a summary and seek employment. Similarly, companies seek candidates based graduate profile of universities to also consider other aspects

[1] e- Qualifications Framework, available online at http://www.ecompetences.eu/

such as additional training and experience [Dorn et al., 2007].

The search for suitable candidates for job openings, or careers covering work places, has always been a complex task, mainly due to different interpretations that each actor has about skills, and hence its many forms representing [Fazel-Zarandi and Fox, 2009]. For example, the term competence in the work context can be used sometimes to refer to actions and their consequences, sometimes as cognitive abilities and personal characteristics [Stevens, 2013]; whereas in academia competencies are expressed in terms of qualifications and certifications (such as diplomas) [Malzahn et al., 2013], or learning outcomes within education processes [Paquette, 2007].

On the other hand, the profiles clearly not describe the competencies or the competency elements, so we can not make a comparison between them [Paquette et al., 2012]. Job offers meet job characteristics such as activities or roles, rather than skills [Bizer et al., 2005]. Moreover, the curriculum of a university degree generally has dependencies between courses, and a general description of the learning objectives to be met to achieve a score [Dorn and Pichlmair, 2007].

Therefore, we need to establish a common environment to achieve a comparison. Through similarity measures and a description of the context around the notion of competencies, we can define a scheme which covers aspects such as extraction, comparison and prediction of competencies. This proposal pursues this objective, which will be explained in the following sections.

3 Architecture

The figure 1 shows the architecture of our proposal, in which the academic and job profiles get into an iterative pipeline process comprising 4 phases: characterization, extraction, comparison and updating.

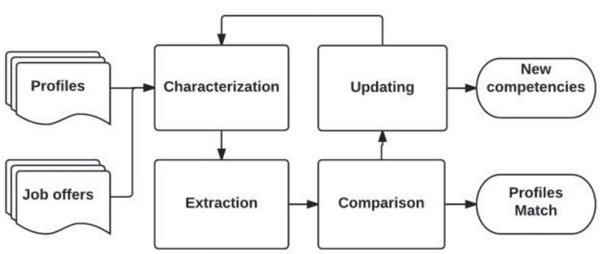

Figure 1: System's architecture

3.1 Characterization

In this phase we propose a scheme based on standard patterns, taxonomies and logical descriptions, which allows the recognition of the competency descriptor elements such as skills and knowledge, to achieve a first scheme of characterization. The characterization phase is based on linguistic patterns, semantic rules and indicators of similarity / dissimilarity, which are the basic input for the following phases.

3.2 Extraction

The extraction phase applies the scheme characterization over the documents. To develop this, low level NLP is combined with the linguistic patterns and semantic rules (defined in the preceding stage) in order to label text pieces in university and job profiles. This phase provides text segments (patterns) which will be compared to determine their similarity.

3.3 Comparison

At this stage we establish measures of similarity between skill, knowledge and competency patterns (identified in the previous phase), achieving levels of closeness between them. To do this, we combine different similarity measures and clustering techniques [Huang, 2008]. This phase will have two output: the first are the groups (clusters) of similar competencies, and the second comprising the remaining competencies. In addition, a set of similarity/dissimilarity indicators for each group are presented.

3.4 Updating

In this phase we analyse groups of similar and dissimilar competencies to determine the inter/intra cluster relationship, looking for several things, among others, those competencies that represent outliers inside and outside the clusters. The output of this phase is recommendations such as: groups of new competencies, more common skills, new knowledge domains, etc.

In this research we focus on the characterization phase, in order to establish our first analysis schema of profiles.

4 Characterization Module

The figure 2 displays the activities carried out in the characterization phase. First, we defined the axioms underlying the scheme. To do this, we used competency concepts and its elements (skills and knowledge), applied in each domain (academic and professional) to identify common contexts. Then, we suggest logical descriptions to characterize these contexts (patterns). To this end, we assumed that there are many definitions of competency and of its components; therefore, we performed a selection process based on a criterion. This criterion is determined by the purpose of comparison.

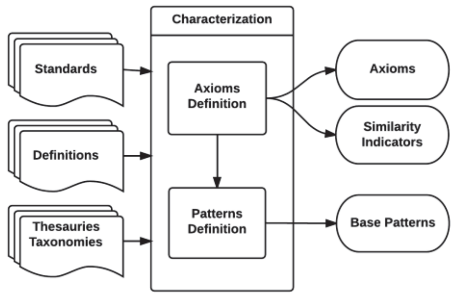

Figure 2: Characterization module

For instance, if the objective is to compare Software Engineering profiles, from South American universities against a set of job offers, the approach includes choosing skill and knowledge concepts, and identify standards, thesauri and taxonomies related to academic and professional context proposed. With them, we would define patterns and initial indicators on Software engineering domain to analyze the academic and work competencies. Thus, we see that the pattern definitions will be determined by generic aspects related to the definition of term competency (axioms), and by issues related to the domain where we want to make the comparison (in our previous example, Software Engineering)

4.1 Axioms definition

The axioms play the role of establishing a generic framework for competency analysis. We carried out the basic definitions of competency elements to consider (knowledge and skills), and used a formal description to make them computable. In the case of university profiles, the axioms that contextualize skills in the academic domain were:

Definition 1: In reference to [Paquette et al., 2012], competency is the ability to use knowledge and skills in work or study situations. According to DISCO competency thesauri [Muller-Riedlhuber, 2009], competencies represent the skills and knowledge of an individual in a specific domain. Taking up the above proposed criteria, examples of these definitions in the domain of software engineering are:

Example 1: Demostrar conocimientos de algorítmica y programación
Example 2: Utilizar los algoritmos de procesamiento de datos para almacenar, acceder y analizar información.

Figure 3: Example of academic competencies

Definition 2: According to DISCO competency thesauri [Muller-Riedlhuber, 2009], skill is the ability to apply knowledge and use know-how to complete tasks and solve problems. In [Paquette, 2007] skills has a taxonomical structured according to the knowledge cognitive level, and can range from generic to specific. For the same example of Software Engineering skills could be:

Example 3: Skills related to the generic skill ***producir***

Generic skill	**Level**	
	1	**2**
Producir	analizar	Deducir, clasificar, predecir, diagnosticar
	sintetizar	Planear, modelar, diseñar

Table 1: Example of skills

Definition 3: Citating DISCO [Muller-Riedlhuber, 2009], knowledge comprises all items related to a field of work or study. According to SWEBOK[2] [Guide in Spanish], knowledge is a hierarchical framework of entities. For the same example of software engineering, knowledge is presented as follows:

Example 4: Knowledge in Software Engineering

Knowledge area	**Knowledge sub-area**
Requerimientos de Software	• Fundamentos de los requisitos • Captura de los requisitos
Diseño de Software	• Estructura y arquitectura de Software • Análisis de la calidad del Software

Table 2: Example of knowledge

For the employment domain, the axioms that characterize the competencies are:

Definition 4: In reference to [Yahiaoui et al., 2006], competency comprises the domain/proficiency levels required in a particular knowledge area. In [Bourque et al., 2003], they affirm that competencies are complemented by the experience, and according to that, competency levels in knowledge areas are suggested. Continuing with the supposed case of Software Engineering domain, competency examples would be:

Example 5: Different proposals of competency levels found in the definitions and standards are:

Proposal	**Competency**
Competence levels [Yahiaoui et al., 2006]	Basic (B o 20%), Application (A o 50%), Master (M o 70%) or Expert (E o 90%).
SW Engineering profiles [Bourque et al., 2003]	NG (New Graduate), G+4 (Graduate with four years of experience), EWSE (Experienced software engineer working in a

[2] SWEBOK Guideline, available online at http://www.swebok.org

	software engineering process group)
Competence proficiency levels (e-CF)[1]	e-1:EQF3 , e-2: EQF 4 and 5, e-3:EQF6, e-4:EQF7 and e-5:EQF8

Table 3: Example of competency

Definition 5: According to DISCO [Muller-Riedlhuber, 2009], skill is the ability and determination to play a role or function. In the context of e-CF[1], the skills are defined in relation to five general action areas which comprise their own abilities. In [Bourque et al., 2003], skills are represented as taxonomic structures that reflect six cognitive levels with their own synonyms, as in the case of Bloom's taxonomy [Capdevila, 2011]. The following example explains this definition:

Example 6: Different skill proposals found in definitions and standards are:

Proposal	Skill
Areas[1]	manage, enable, run, build and plan
Taxonomical levels [Bourque et al., 2003]	knowledge, comprehension, application, analysis, synthesis and evaluation

Table 4: Example of skill

Definition 6: Knowledge is an object of competency with a specific level of competency[1]. According to [Yahiaoui et al., 2006], the objects of competency may be technological topics belonging to specific knowledge area or software artifacts. To Sicilia [Sicilia et al., 2005], an artifact is "an at least partially tangible thing which was intentionally created by a person", on the other hand, competency levels can be defined based on the examples given in the table 3. The following example explains this definition:

Example 7: Artifacts in Software Engineering

Area	Knowledge
Requerimientos de Software	Documentos de requisitos
Diseño de Software	Patrón de diseño, traza de requerimiento, programa

Table 5: Example of knowledge

4.2 Patterns definition

Once raised the axioms, we proceed to define the linguistic patterns for each of them. In the Spanish language, the basic structure of a sentence consists of two parts, subject and predicate, being the core of the subject noun and the verb predicate core. In Computational Linguistics, the subject of the sentence is associated with a noun phrase (Noun phrase), and predicate with a verb phrase (Verb Phrase) [Manning and Sch¨utze, 1999]. In the case of profiles, the expressions used to describe competencies are descriptive sentences, but with the difference that they have a tacit subject. The Figure 4 presents the linguistic analysis of a classical competency sentence: "Gerenciar centros de cómputo".

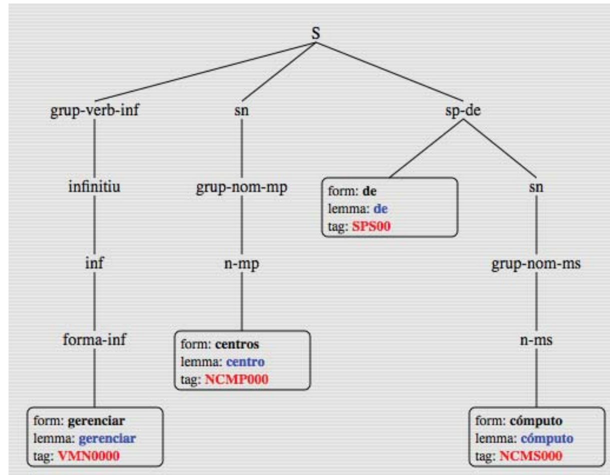

Figure 4: Linguistic analysis of a classic competency sentence in Freeling[3]

Based on the axioms defined in paragraph 4.1 for the academic context, we can say that competency is the union of a verb phrase and noun phrase, where the noun phrase (Noun phrase) represents knowledge and verbal phrase (Verb phrase) represents skills. Table 6 presents the proposed patterns according to the axioms for the academic context, as well as examples found in the profiles.

Element	Pattern	Example
Knowledge	Noun Phrase: (NP) [(NP)(NP)][(NP)(Prep)(NP)]	-Proyecto -Sistema Operativo -Programa de software
Skill	Verb Phrase: (VP)[Flexión Verbal(NP)]	-Diseñar -Gestionar -Gestión
Competency	Noun Phrase + Verb Phrase: (NP) + (VP) [(NP)(NP)][(NP)(Prep)(NP)] + (VP)[Flexión Verbal(NP)]	-Diseñar + programas de software -Gestión + de sistemas operativos

Table 6: Academic competency patterns

Similarly, we can define a competency characterization to occupational context (see Table 7), where competences are defined as noun phrases (Noun phrase). Furthermore, the

[3] Freeling, available online at http://nlp.lsi.upc.edu/freeling/demo/demo.php

skills are presented as the combination of verbal and noun phrases, while knowledge is formed based on noun phrases, as in the academic context.

Element	Pattern	Example
Knowledge	Noun Phrase: (NP)|[(NP)(NP)] | [(NP)(Prep)(NP)]	-Java -Sistemas Operativos -Oracle -Bases de Datos -Patrones de diseño
Skill	Verb Phrase + Noun Phrase: (VP) + (NP) (VP)[Flexión Verbal(NP)] + [(NP)(NP)][(NP)(Prep)(NP)]	-Gerenciar + redes de computadoras -Gestión + de servidores Linux
Competency	Noun Phrase: (NP) (NP) | [(NP)(NP)] | [(NP)(Prep)(NP)]	-Desarrollador Máster -Director de Proyectos Junior

Table 7: Occupational competency patterns

5 Experimentation

5.1 First Experiment: Text analysis

With the purpose to analyze the scheme patterns in both contexts, we make a first approximation of its uses in the profiles taken from universities and industrial contexts. The Figure 5 shows an example of our characterization's definitions in pieces of competency profiles, taken from universities and work contexts, where we see knowledge patterns highlighted in blue and skill patterns highlighted in red.

University profile	***Título:*** *Ingeniero en Software* ***Función:*** **Diseñar, implementar y evaluar componentes, programas y sistemas.**
Job Offer	***Cargo:*** **Desarrollador de Software** ***Descripción:*** Ingeniero de Software especializado en **programación de sistemas**. El candidato deberá tener la habilidad de **priorizar** actividades para **alcanzar** fechas de entrega. (Es deseable que tenga habilidades en la **gestión de proyectos**)

Figure 5: Pattern annotation over profiles

Since in both contexts competencies are interpreted differently, we can note difficulties in connecting skills in the texts. At first glance, we show that university profiles contain complex sentences with several words tagged as skills and knowledge. On the other hand, job offers contain a greater presence of knowledge descriptions, although some of them can be understood as skills, transforming these words in its verbal inflection (Programming → To program). In conclusion, there is a high degree of ambiguity in the text, which does not fully identify competencies or their components.

From this input, the comparison phase of our system may establish metrics, strategies to compare the information identified in the analysis of texts of competencies, obtained by our patterns, as the case that we come to show.

5.2 Second experiment: standards-based analysis

The Figures 6 and 7 show an example of alignment of the profiles presented in the first experiment (see figure 5), through two standards related to Computer Sciences domain. Figure 6 compares the competency of the university profile: "Diseñar, implementar y evaluar componentes y programas de sistemas" (Design, Implement and evaluate-system program components) with two standards: DISCO [M ̈uller-Riedlhuber, 2009] and SWEBOK [Sicilia et al., 2005].

To this end, we used the characterization patterns defined in section 4.2, and an iterative algorithm to compare the patterns present in the sentence against the lower levels of the hierarchy of each body of knowledge. We selected a candidate pattern based on pattern matching hits, which is now compared against the lower levels of the hierarchy to find a match. For this experiment, the pattern (system program) becomes the candidate pattern. Figure 6 shows the result of the iteration, where the path through the hierarchies gets some indicators such as: domains of knowledge, and the competence levels related to the pattern.

PATTERN	DISCO	SWEBOK
[system program]	Computing (1st Level) – Programming (2nd Level) **Term:** **System Programming** **Phrase** • create program modules and procedures	**KA:** Software Construction (1st Level) – **Breakdown topic** (2nd Level) Practical Considerations – **Construction languages** (3rd Level) – **Topic** **System Programming Notations**

Figure 6: Tracking academic patterns on DISCO and SWEBOK

We can see that there are ambiguities in text segments of profiles, clearly demonstrated by the different topics that form the comparison route (systems software, systems programming, systems programming notation). These three topics reflect different domains of knowledge in relation to the original pattern, although there is a certain similarity between them.

We repeated this analysis on job offer patterns. Figure 7 shows the result of the experiment. After several iterations, the candidate pattern was "programación de sistemas" (system programming). In this case, the comparison route through the hierarchy reveals a relative similarity between the candidate pattern in figure 7 (system program) and the second level of DISCO (Computing/Programming), and the third level of SWEBOK (Practical considerations/Construction Languages).

PATTERN	DISCO	SWEBOK
[software programing] [system programming]	Computing (1st Level) – Programming (2nd Level) –IT Consultancy (3rd Level) – SW. Evaluation – SW Design Term: System Programmes Phrase • analyse the efficiency of applied software programmes • perform tests on established software programmes	KA: Software Construction (1st Level) – Breakdown topic (2nd Level) Practical Considerations – Construction languages (3rd Level) – Topic System Programming Notations

Figure 7: Tracking occupational patterns on DISCO and SWEBOK

The results of this second set of experiments, gives us a first idea of the level of similarity between academic profile and the job offer profile. Therefore, we suppose that it is possible to use the standards as a means of comparing the profiles in both contexts [Rudzajs and Kirikova, 2011], thereby establishing a midpoint that solves the problem of ambiguity. The comparison phase has on these results an important input. For that, we need to refine concepts like "relative similarity", used on the comparison phase, based on the linguistic patterns defined in this paper.

6 Conclusions

Our proposal of a characterization module based on a modeling approach, allows that the Profile Management Architecture will be adaptable to the context. The definition of axioms and linguistic patterns for competencies and its basic elements, contributes to the development of the other modules of the architecture. For example, facilitates the process of pattern extraction (extraction phase), the matching of competencies (comparison phase), and also provide a framework for the upgrade of the competencies (updating phase).

The experiments lead us to think that the patterns proposed in the characterization module allow us the identification of knowledge and skill within documents, and provide a first approach towards defining a process for comparing the profiles. This approach will be evaluated in the comparison module through mechanisms of matching of profiles, using as input the patterns identified in the texts using our axioms and linguistic patterns.

The application of low-level linguistic patterns on the corpus of profiles, gives us the opportunity to experiment with machine learning algorithms under a semi-supervised approach, especially in the step of updating of competencies (learning skills, knowledge and competencies).

In order to improve this research, future works will be directed towards developing of the modules of extraction and comparison of our architecture, the formal definition of axioms and linguistic patterns using descriptive logic, and the implementation of machine learning algorithms for the phase of competency upgrading.

Acknowledgments

Dr Aguilar has been partially supported by the Prometeo Project of the Ministry of Higher Education, Science, Technology and Innovation of the Republic of Ecuador.

References

[Bizer et al., 2005] Christian Bizer, Ralf Heese, Malgorzata Mochol, Radoslaw Oldakowski, Robert Tolksdorf, and Rainer Eckstein. The impact of semantic web technologies on job recruitment processes. In Wirtschaftsinfor-matik 2005, pages 1367–1381. Springer, 2005.

[Bourque et al., 2003] Pierre Bourque, Luigi Buglione, Alain Abran, and Alain April. Bloom's taxonomy levels for three software engineer profiles. In Software Technology and Engineering Practice, 2003. Eleventh Annual International Workshop on, pages 123–129. IEEE, 2003.

[Capdevila, 2011] Javier P´erez Capdevila. Desde los mapas de conocimientos hacia los mapas de competencias. Avanzada Cient´ıfica, 13(1):67–75, 2011.

[Dorn and Pichlmair, 2007] Jrgen Dorn and Markus Pichlmair. A competence management system for universities. In ECIS, pages 759–770, 2007.

[Dorn et al., 2007] J¨urgen Dorn, Tabbasum Naz, and Markus Pichlmair. Ontology development for human resource management. In Proceedings of the 4th International Conference on Knowledge Managements, pages 109–120. Citeseer, 2007.

[Draganidis and Mentzas, 2006] Fotis Draganidis and Gregoris Mentzas. Competency based management: a review of systems and approaches. Information Management & Computer Security, 14(1):51–64, 2006.

[Fazel-Zarandi and Fox, 2009] Maryam Fazel-Zarandi and Mark S Fox. Semantic matchmaking for job recruitment: an ontology-based hybrid approach. In Proceedings of the 8th International Semantic Web Conference, 2009.

[Fazel-Zarandi, 2013] Maryam Fazel-Zarandi. Representing and Reasoning about Skills and Competencies over Time. PhD thesis, University of Toronto, 2013.

[Huang, 2008] Anna Huang. Similarity measures for text document clustering. Proceedings of the sixth new zealand computer science research student conference (NZCSRSC2008), Christchurch, New Zealand. 2008.

[Lindgren et al., 2004] Rikard Lindgren, Ola Henfridsson, and Ulrike Schultze. Design principles for competence management systems: a synthesis of an action research study. Mis Quarterly, pages 435–472, 2004.

[Malzahn et al., 2013] Nils Malzahn, Sabrina Ziebarth, and H Ulrich Hoppe. Semi-automatic creation and exploitation of competence ontologies for trend aware profiling, matching and planning. Knowledge Management & E-Learning: An International Journal (KM&EL), 5(1):84–103, 2013.

[Manning and Sch¨utze, 1999] Christopher D Manning and Hinrich Sch¨utze. Foundations of statistical natural language processing. MIT press, 1999.

[Mena et al., 2007] Elena Mena, Mar´ıa Pareja, Miguel-Angel Sicilia, and Elena Garc´ıa-Barriocanal. Deriving competencies from the swebok ontology. In 2nd International Workshop on Ontology, Conceptualization and Epistemology for Software and System Engineering, 2007.

[M¨uller-Riedlhuber, 2009] Heidemarie M¨uller-Riedlhuber. The european dictionary of skills and competencies (disco): an example of usage scenarios for ontologies. In I-SEMANTICS, pages 467–479, 2009.

[Paquette et al., 2012] Gilbert Paquette, Delia Rogozan, and Olga Marino. Competency comparison relations for recommendation in technology enhanced learning scenarios. In Proceedings of the 2nd Workshop on Recommender Systems for Technology Enhanced Learning (RecSysTEL 2012), volume 896, pages 23–34. Citeseer, 2012.

[Paquette, 2007] Gilbert Paquette. An ontology and a software framework for competency modeling and management. Educational Technology & Society, 10(3):1–21, 2007.

[Poblete, 2006] Manuel Poblete. Las competencias, instrumento para un cambio de paradigma. 2006.[Rieu et al., 2005] Christine Rieu, C´ecile Million Rousseau, and C Roche. Gestion des comp´etences: un mod`ele op´erationnel a ` base dontologie. In Colloque Du E-Management a ` la E-RH, 2005.

[Rudzajs and Kirikova, 2011] Peteris Rudzajs and Marite Kirikova. Mediated competency comparison between job descriptions and university courses. Scientific Journal of Riga Technical University. Computer Sciences, 44(1):48– 56, 2011.

[Schmidt and Kunzmann, 2006] Andreas Schmidt and Christine Kunzmann. Towards a human resource development ontology for combining competence management and technology-enhanced workplace learning. In On the Move to Meaningful Internet Systems 2006: OTM 2006 Workshops, pages 1078–1087. Springer, 2006.

[Sicilia et al., 2005] Miguel- Angel Sicilia, Juan J Cuadrado, and Daniel Rodr´ıguez. Ontologies of software artifacts and activities: Resource annotation and application to learning technologies. In SEKE, pages 145–150, 2005.

[Stevens, 2013] Gregory W Stevens. A critical review of the science and practice of competency modeling. Human Resource Development Review, 12(1):86–107, 2013.

[Yahiaoui et al., 2006] Leila Yahiaoui, Zizette Boufa¨ıda, and Yannick Pri´e. Semantic annotation of documents applied to e-recruitment. In SWAP, 2006.