POLITECNICO DI TORINO

SCUOLA DI DOTTORATO

Dottorato in Ingegneria Informatica e dei Sistemi – XXV ciclo

Tesi di Dottorato

Exploiting semantics to support education and labor worlds

Valentina Gatteschi
matricola: 169464

Tutore
prof. Claudio Demartini

Coordinatore del corso di dottorato
prof. Pietro Laface

Marzo 2013

Contents

List of Tables

List of Figures

Chapter 1

Introduction

The last decade has been characterized by a marked increase in mobility across Europe. In fact, since the EU enlargement of 2004, that removed the barriers with the Eastern countries, the migration in Europe tripled, from about 1.6 million in 2003 to about 4.8 million in 2009 [NIESR, 2011]. The reasons behind this growth are linked both to the search for new employment opportunities, as in the case of East-West migration flows, both to the need to acquire additional competences in a foreign country.

Nevertheless, mobility is still threatened by cultural, linguistic and systemic differences. In fact, as a matter of example, students willing to continue their study career abroad, encounter some difficulties when choosing courses to attend in order to acquire missing competences, due to the fact that qualifications are usually articulated in different ways, in which “contents” (in terms of learning outputs, or outcomes) are described heterogeneously; moreover, qualification pillars (like knowledge, competences, skills, etc.) assume different meanings in the specific national domain, with serious consequences on mutual understanding. In a similar way, in the human resources acquisition context, when recruiters have to match required competences with the ones possessed by a job seeker, a simple check on owned qualifications could not be sufficient, since the same qualification, in different countries, could provide different competences, with different levels of detail.

In order to overcome the above limitations, the European Commission proposed, in 1987, the Erasmus (European Community Action Scheme for the Mobility of University Students) Programme, an exchange programme[1] targeted to Higher Education students.

However, if student transfers between Universities are a praxis, in the Vocational Educational and Training (VET) this harmonization process is still in progress.

[1]http://ec.europa.eu/education/lifelong-learning-programme/erasmus_en.htm

In fact, strategies adopted to overcome this gap in qualifications readability, usability and comparison, so as to support the development of “a knowledge-based Europe” and to ensure that “the European labor market is open to all” - as it is expected by the Bruges-Copenhagen process[2] - have been presented less than a decade ago, in the Maastricht Treaty[3]; specifically, the Treaty stated that the efforts in the life-long learning perspective had to be focused on the “development of an open and flexible European qualifications framework, founded on transparency and mutual trust”, by also underlying the need to develop a European credit transfer system for VET based on competences and learning outcomes. On April 23, 2008, the European Parliament Council took a step forward in this direction, by defining the eight levels of the novel European Qualification Framework (EQF)[4] instrument, and by also identifying precisely the semantic - among others - of qualification, learning outcome, knowledge, skill, and competence concepts, thus enhancing the creation of the expected shared understanding in the life-long learning domain. Moreover, still in the transparency direction, according to the above Recommendation, Member States were also encouraged to reference their national qualifications (systems, or frameworks) to the European instrument, thus favoring the necessary harmonization process.

The harmonization process is still under development in almost all the Member States. Moreover, in this view, several experiences devoted at supporting mobility by increasing transparency and mutual trust have been recently carried out. This thesis aims at presenting Ph.D. research activities performed within two of them, namely the TIPTOE “*Testing and Implementing EQF and ECVET Principles in Trade Organizations and Education*” project[5] and the MATCH “*Informal and non-formal competences matching device for migrants’ employability and active citizenship*” project[6].

The goal of both projects was the exploitation of European instruments and technology, on the one hand, in order to support the construction and comparison of qualifications (the TIPTOE project), on the other hand, with the aim of connecting migrants’ competences acquired in formal, non formal and informal contexts to job offers, in a job matchmaking scenario (the MATCH project).

In this view, two platforms have been developed:

- the TIPTOE platform, collecting occupational and educational profiles of the

[2]http://ec.europa.eu/education/pdf/doc125_en.pdf

[3]http://ec.europa.eu/education/news/ip/docs/maastricht_com_en.pdf

[4]http://ec.europa.eu/education/lifelong-learning-policy/eqf_en.htm

[5]http://www.evta.net/tiptoe/home_tiptoe/

[6]http://match.cpv.org/

retail sector, enabling end-users to find commonalities and differences among them, with the aim of easing the creation of a “common” profile based on labor market requirements and training outputs;

- the LO-MATCH platform, enabling the insertion of EQF-based job offers and demands and the computation of the match among them.

In order to overcome the marked differences in the way qualifications, curricula and job offers are expressed, semantic instruments have been exploited.

In particular, during the development of the TIPTOE platform, an ad-hoc taxonomy structuring concepts belonging to occupational and educational descriptions has been created. Composing elements (the learning outcomes) have then been annotated by making reference to the above concepts. Finally, four different approaches for computing the similarity between elements have been devised. Even though the exploitation of such a tool reduced the workload of project partners during the creation of the common profile - a task that requires a considerable amount of time, when manually performed - a huge effort has been spent in the creation of the taxonomy and in the annotation of learning outcomes.

The lesson learnt within the TIPTOE project has been considered during the creation of the LO-MATCH platform: here an existing semantic thesaurus (WordNet[7]) has been exploited for (semi-)automatically annotating acquirements and requirements. Then, a facilitator, allowing job seekers and recruiters to add learning outcomes either by browsing or by performing a free text search on pre-inserted profiles has been created. Finally, a way to compute the match between job offer and demand has been formulated.

The present thesis is organized as follows: Chapter 2 presents an overview of research works exploiting semantic technologies both in the Education and Training scenario (Section 2.3), both in a job matchmaking context (Section 2.4).

In Chapter 3, the overall methodology adopted during the TIPTOE project is introduced (Sections 3.1, 3.2, 3.3), together with a set of additional services targeted to end-users (Section 3.4).

Chapter 4 is about the MATCH project (Sections 4.1 and 4.2), the development of the LO-MATCH platform (Sections 4.3 and 4.4) as well as its functionalities and interfaces (Section 4.5), and additional features using a tag-cloud based representation of job offers and demands (Section 4.6).

Finally, conclusions are drawn in Chapter 5.

[7] http://wordnet.princeton.edu

Chapter 2

Opportunities offered by semantic technologies

2.1 Introduction

From a lexical point of view, the word semantics, from the Greek *sēmantikós*, refers to the study of the meaning of words, phrases, sentences and large bulk of text. Hence, semantics is often associated with various fields of study like, among others, philosophy, philology, communication, and semiotics. From the perspective of the information and communication sciences and, in particular, of the Internet, the concept of semantics started to be extensively used in 2001, when the term Semantic Web was coined to refer to a technological revolution aimed at transforming the classic web pages into transparent information sources to be read and understood by machines [Berners-Lee et al. 2001]. In order to support the vision of the Semantic Web, a way had to be defined to precisely identify data within online (hyper-)documents. Moreover, a means for letting computer agents perform automatic processing over data by mimicking human reasoning processes had to be found as well. The above goals were achieved thanks to efforts of the World Wide Web Consortium (W3C)[1], which defined a number of instruments that represented the ground for the evolution of web 1.0 into its recent shapes. In particular, a means for keeping separate in web pages data presentation from data themselves was developed. This way, it became possible to make it explicit the informative content of documents on the web. Then, taking inspiration from human mental processes, a number of solutions for creating an articulated data model to be possibly exploited for enriching the above information with a meaning were defined. In such a model, meaning is expressed by making reference to concepts, which are in turn linked to

[1] www.w3.org

other concepts through a complex network of relations. Finally, web content producers and consumers were offered the opportunity to publish and exchange information by annotating it, i.e., by associating it with identifiable resources belonging to the knowledge model above. This way, machines were opened the possibility to read and elaborate information that was originally meant for being accessed and processed just by humans.

By leveraging on the above possibility, several software solutions started to be developed to exploit semantic data for dealing with knowledge-related problems. In particular, semantic-based approaches and related technologies were initially applied in intensive knowledge management scenarios like, for instance, bioinformatics and health sciences in general. Then, they were exploited for drafting the standards for data exchange and integration as well as for service interoperation over the Internet [Yimam-seid and Kobsa, 2003]. They were also chosen as the core building block for the realization of sophisticated distributed applications, including those referred to as fostering social and networking relations. In the last years, their helpfulness in supporting humans in performing a huge number of time consuming tasks requesting to filter, compare, aggregate and evaluate heterogeneously structured information has been demonstrated in many specific application fields.

Very recently, after a number of laboratory experiments, semantic techniques have been applied also to the probably more obvious domain (though more irksome, due to its general purposeness) represented by web search. Today, semantic search is associated with the idea of an information retrieval process that goes beyond a pure literal-lexical match based on keywords. Instead, semantic search is regarded as being capable of exploiting the meaning of words and sentences and of considering individual terms in their surrounding context in order to anticipate users' needs and provide them with the best results possible by simply asking them to tell what they want.

In fact, traditional search strategies where a number of keywords need be combined to generate the query string could sometimes fail in returning results expected, e.g., because of the order of terms or of logical operators used, because of the ambiguous meaning of words, etc. An example could be represented by a search for an hotel in Rome. A possible query string could be hotel Roma. When searched with the above keywords, Google would return in the top ranked results two pages actually making reference to hotels in the city of Rome (Figure 2.1).

However, the query would return at the same time the web pages of other hotels named Roma but located in different cities (Figure 2.2).

Another example could be represented by situations where the same term is used with different meanings, as shown in the tweets about light (Figure 2.3).

A semantic engine could try to mimic human reasoning to figure out the exact meaning of the terms concerned by exploiting context cues, i.e., by considering each term with the surrounding words in the same sentence. Similarly, in some cases it

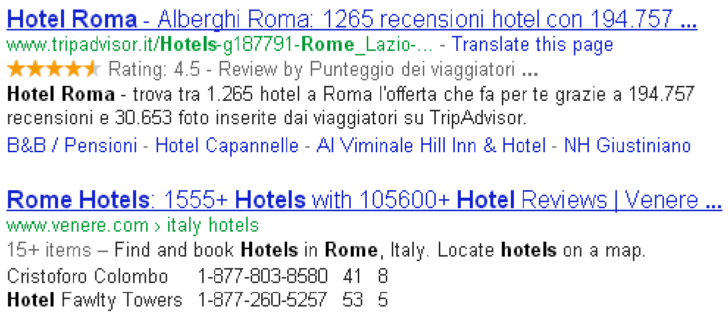

Figure 2.1. Hotels in Rome

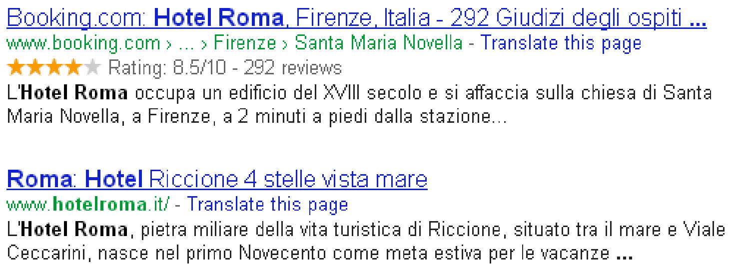

Figure 2.2. Hotels named “Roma” but located in different cities

might be helpful to extend the search space identified by user’s query by taking into account terms conceptually close to the ones that have been explicitly typed[2]. For

[2]www.google.com/insidesearch/features/search/knowledge.html

Figure 2.3. Tweets containing the word “light”

instance, when searching for *money*, proximity relations could be exploited and other concepts that could be of interest might be also considered, like *finance*, *commerce*, *bank*, *market*, but also *buying*, *selling*, *exchanging*, etc.

From this viewpoint, the education and learning field, as well as job seeking and job recruiting could benefit from the exploitation of semantic instruments.

In fact, even though in Europe during the last years several instruments to support mobility and employability have been defined (e.g., "EQF European Qualification Framework"[3] and ECVET "European Credit System for Vocational Education and Training"[4]), students and workers who decide to spend a working period abroad still encounter several difficulties in the recognition/validation of their qualifications, which can be mainly associated to information asymmetries among students, job seekers, employers and training centers: this is due to the fact that, usually, for a student who decides to continue his/her study career abroad it's difficult to find (within courses descriptions) competences he/she need to achieve, or classes he/she needs to attend to obtain a given qualification.

This difficulty arises from the heterogeneity of qualifications structure, and from the lack for well-established definitions: in fact, as a matter of example, two countries may show qualifications articulated in different ways, in which contents could be described heterogeneously; moreover, the fact that knowledge, competences, skills concepts could assume different meanings in the specific national domain seriously threatens mutual understanding.

Similar considerations apply for the working dimension: in fact, frequently, human resources staff (which have a huge psychology background, but are possibly lacking in technical knowledge), when examining a curriculum e.g. in digital form, often perform a sort of "keyword-based" analysis on the basis of requirements indicated by the business department that is looking for new workers to hire; however, when specialized competences required to fully understand the curriculum are not available, a possibly adequate match between demand and offer could be lost, e.g. because of the use of non-aligned vocabularies.

In this Chapter, an investigation on how semantic technologies have been used for supporting mobility and employability is presented. In particular, first, an overview of semantic instruments is presented in Section 2.2. Then, Section 2.3 analyses how semantics has been used to foster students' mobility and to support the development of training courses from three perspectives, namely *the creation of the knowledge base*, *the integration of heterogeneous systems and definition of inference rules*, and *the visualization of training courses*. Finally, Section 2.4 investigates the state of the art in the exploitation of semantic technologies to support job matchmaking.

[3] http://ec.europa.eu/education/lifelong-learning-policy/eqf_en.htm

[4] http://ec.europa.eu/education/lifelong-learning-policy/ecvet_en.htm

In particular, research works are presented by making reference to the three dimensions characterizing the creation of a semantic job matchmaking system: *knowledge representation*, *the annotation* and *the computation of the match*. In this Section, an overview of commercial platform supporting job seekers and recruiters is proposed (Section 2.4.2), together with some concluding remarks (Section 2.4.3).

2.2 The world of semantics

In the computer science domain, the concept of semantics and Semantic Web are strictly related to terms like knowledge representation and knowledge management. In fact, in the vision of its founder Tim Berners-Lee[5], the goal of the Semantic Web was about building a framework able to support the creation of knowledge-based services and applications, by leveraging on a shift from *a web of documents* to *a web of data*. In this vision, the links are between ad hoc data representations, or metadata, rather than between online documents, or web pages, reporting data themselves. Therefore, besides serving as a way for splitting information from presentation, the contribution of Semantic Web tools and techniques is in the direction of fostering the extraction and, then, the practical usage of knowledge embedded into data.

A key role towards the implementation of the above vision was played by the introduction of enabling technologies like the Extensible Markup Language (XML)[6], the Uniform Resource Identifier (URI)[7], the Resource Description Framework (RDF)[8] and the Web Ontology Language (OWL)[9]. In fact, by using the XML, information originally blended with presentation into HTML pages can be easily made explicit. Then, the RDF can be used as a data model for enriching information with a meaning, that can be leveraged while publishing and exchanging information on the web by making reference to individual resources identified by their URI. Finally, with OWL, the above meaning can be associated with concepts, linked to a knowledge base where relations between concepts describing such a meaning can be formalized.

As demonstrated by the technological focus, it is first of all a matter of representing knowledge. The concept of knowledge is, however, quite abstract. In the field of semantics, knowledge is regarded as the subset of information that can be described in an explicit and formal way. As a matter of example, knowledge could be expressed by means of sentences like *the lion is a carnivore* or *a carnivore is an*

[5]http://www.w3.org/People/Berners-Lee/

[6]http://www.w3.org/XML/

[7]http://www.ietf.org/rfc/rfc3986.txt

[8]http://www.w3.org/RDF/

[9]http://www.w3.org/OWL/

animal, where concepts like lion, carnivore and animal are used. However, in the perspective of enabling machines to understand and exploit the above knowledge, it is essential to identify suitable mechanisms for framing these concepts into structured models, where the underlying meaning and relations are elicited and, hence, made processable.

Common implementations of such models in the field of computer science are represented by taxonomies and ontologies, where meanings and relations associated with concepts of interest are elicited for a specific domain.

A taxonomy is the result of the process of identifying, grouping and naming individuals on the basis of shared characteristics. In a taxonomy, groups are organized in a structure encompassing super and sub-groups, contributing at generating a classification. An ontology is defined as *a formal, explicit specification of a shared conceptualization* [Gruber, 1993a]. A *conceptualization* refers to an abstract model of some phenomena in the world able to identify the relevant concepts involved. The term *formal* refers to the fact that the ontology has to be read by machines. The word *explicit* is used to express the fact that concepts used are defined in an explicit way. Finally, an ontology is said to be shared, since concepts used have to be chosen so that they are able to describe a consensual knowledge which will be exploited by various actors. Thus, taxonomies and ontologies are central to the Semantic Web, since they represent a *common vocabulary* that allow machines to agree upon terms to be used for communicating [Guarino, 1998].

Though the definitions of taxonomies and ontologies could be rather abstract, when it comes to considering their role in the field of semantics and practical examples are taken into account, things become rather clearer. For instance, the excerpts of a possible taxonomy (left) and ontology (right) for a particular domain are illustrated in Figure 2.4. The subject is still the animal world. As it can be seen, the taxonomy is basically represented by a tree, with upper (more general) groups incorporating lower (more specific) elements, or concepts. In the ontology, the basic subsumption links are enriched with new relations. That is, besides *is-a* links, new relations making reference to the fact that a carnivore *eats* animals whereas an herbivore eats plants are elicited. Moreover, plants can be even related to carnivores and, specifically, to one particular kind of carnivore like the lion that, for instance, *hides in* the long grass (that is a type of plant) for hunting preys. This way, a kind of graph (directed, in this case) is created.

Once the issue of representing knowledge has been dealt with, the second step is about how to make use of it. In most of the cases, the goal of knowledge-based systems is to support humans' work by automating knowledge intensive tasks, which generally imply to use some human-like reasoning mechanism onto certain domain knowledge to address a particular problem. Depending on the particular reasoning rules being exploited and the specific application scenario being considered (classification, monitoring, prediction, planning, comparison, etc.), the goal could be

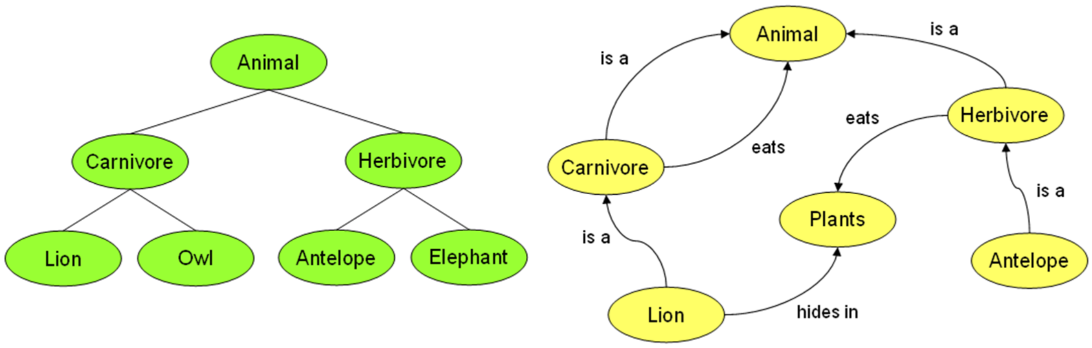

Figure 2.4. Example of taxonomies and ontologies

to deduct implicit information, infer logical relations, resolve heterogeneities, draw conclusions, etc.: in one word, generating new knowledge [Guarino, 1998].

What is really impressive is how much the goal of automatic reasoning could be simplified by working on the formal models discussed above. In fact, starting from concepts organized in trees and graphs which are usually developed to formalize a particular domain, semantic engines are provided with much more information to work with than with simple keywords and effective inference rules can be produced by letting a machine reason using concepts that it knows are conceptually related. To make an example, on the one side the taxonomy above is expressing the fact that a carnivore is a sub-type of animal, so that any carnivore is an animal (but not every animal is a carnivore). On the other side, the ontology is saying that, since lions eat other animals, herbivores like antelopes can be possibly eaten by carnivores like lions. In such a scenario, when exploited for finding matches, for instance, with the word elephant, a semantic tool like a search engine would return hits with both antelope (which, in a possibly extended model, should be expected to be a sibling of elephant) as well as with herbivore and animal (parent concepts) and with plants (because of the eating relation), of course with different scores.

Now, assuming that the job seeking and job recruiting domains have been modelled in the above terms, translating the envisioned reasoning methods to address the problem of finding the best match between job offers and demands appears straightforward. For instance, a really basic way for exploiting the existence of formal knowledge models to support job seekers' and recruiters' tasks would be to let a possible matching engine neglect intentional or unintentional variations in terms mentioned in users' search strings (plurals, misspelled items, abbreviations, etc.), thus improving system interaction and usability. A smarter way for using the above domain models could be to let the semantic engine extract the meaning associated with a given sentence used to describe job seeker's achievements or job posting requirements. To this aim, the conceptual distance in the models and the physical distance in the sentence between concepts and terms could be used together with relations between words to improve the engine ability to identify similarities between résumé and job offer descriptions. Finally, new knowledge could be inferred, by searching for what was not explicitly mentioned in the résumé or in the job offer themselves (because the form used was free, or limited, or not developed with the goal of supporting job matchmaking, like in the case of social platform profiles).

Similarly, when the focus is on mobility and learning, students' acquirements could be modeled as described above, and matched against training courses, both in order to find those training paths providing missing competences, both for supporting the validation of prior learning.

Although, as shown, the idea behind semantic processing is quite trivial and opportunities seem enormous, it is worth saying that the real advantages coming from the exploitation of such techniques become evident only once information has been properly annotated with respect to a comprehensive data model for the domain and user-friendly reasoning techniques tailored to the particular application problem being addressed have been developed. Unfortunately, as it will be shown in the following, managing in the broad sense of the term (i.e., creating, exploiting, and maintaining, etc.) such knowledge system is all but trivial.

2.3 Semantics and Education

In the last decades, the way people approach learning has changed. In fact, nowadays, more and more students participate in forums, or surf the internet in order to gain knowledge or find resources for their homework: they have learnt how to search, reuse and (often) publish information.

A recent survey [Netcraft, 2011] identified, in February 2013, 630.795.511 websites that are accessible on the Internet. Two years before, during the same period, the number of active websites was only 284.842.077. Even though some people argue that "The more information the better!", identifying searched information is sometimes difficult; hence, available data should somehow be processed, to simplify teachers' and students' searches.

A first study [Ohler, 2008] investigated how semantic technologies could be exploited in order to support learners, and identified three areas of impact: knowledge construction, personal learning network maintenance and personal educational administration. According to the author, semantic instruments could be used a) to produce, in reply to a search, a multimedia report drawn from many sources such as websites, chapters in textbooks, speeches posted on YouTube, etc., rather than a list of hits (knowledge construction area); b) to identify relevant information from blogs, podcasts and other semantically accessible sources and to provide an information synthesis tailored to the student's personal learning objectives, thus creating personal learning networks mainly built around subjects, rather than around services (personal learning network maintenance); c) to compare courses provided by different institutions in order to improve students mobility (personal educational administration).

Other authors [Tiropanis et al. 2009] carried out a survey of semantic tools and services relevant to UK higher education and identified learning and teaching challenges to which they could be relevant. In particular, the working group involved in this research activity recognized that main areas that could benefit from semantic technologies are a) the courses creation and revision, b) the recommendation of resources matching the topics of students' assignments, c) the creation of groups for collaborative work on the basis of learners' background, personal preferences and successful prior collaboration, d) the creation of links between discussions for enhancing critical thinking, e) the match of people and re-sources across schools of the same or different institutions, as a support to cross-curricular activities, f) the creation of personalized knowledge, g) the support to group knowledge construction.

Starting from the above challenges, the authors tested available tools and services and identified best practices for four main categories. In particular, for the first category collaborative authoring and annotation tools, 7 services have been identified, for the second one, searching and matching tools using semantic technologies, 2 tools have been found, for the third category, repositories for import/export of data using semantic technologies, 5 tools have been detected, whereas the fourth category, infrastructural tools and services for the integration of data sources across organizations in interoperable semantic formats, includes 4 services. Hence, while several tools for collaborative authoring and annotation already exist, only few instruments for performing searches within institutions have been developed.

Usually, people decide to spend a studying or working period abroad to acquire missing competences, or to find better working opportunities. In this context, students need to identify the set of competences that could actually be gained abroad. Unfortunately, besides language barriers, other difficulties arise since, usually, countries present different training systems, and it is extremely difficult to identify competences that are provided by a qualification, or are held by workers.

The first instrument that supports European mobility is the European Qualification Framework, a framework classifying qualifications according to 8 levels that identifies precisely the semantic – among others – of qualification, learning outcome, knowledge, skill, and competence concepts, thus enhancing the creation of the expected shared understanding in the life-long learning domain. Thanks to EQF, training systems (or workers) defining their courses (or curricula) according to the above guidelines have more possibilities of being understood by the other actors. Nevertheless, due to the huge amount of data, a mere application of EQF principles carried out on manual basis, in most of the cases, is not sufficient. Hence, ad-hoc instruments addressing the following issues, should be created:

- concerning *students' mobility*, EQF-based matchmaking systems that allow students to analyse qualifications that are provided by other training institutions and to identify classes that they need to attend, to fill their gaps, must be developed;
- with regard to the *quality of the qualification offer*, instruments that enable training authorities to develop qualifications that answer to the needs of the labor market and that could combine training modules by taking into account their level of difficulty, as well as relations among them, should be created;

These instruments should be capable of comparing heterogeneous training systems, curricula and job offers and should allow the user to perform not only keyword-based searches but also more complex queries that are based on relations among concepts; hence, they should exploit semantic instruments, such as taxonomies or ontologies.

Moreover, the above tools should a) rely on a knowledge base that collects relevant information or on different merged knowledge bases, because of the heterogeneity of the information to be processed; b) foresee the definition of inference rules and perform matchmaking, to provide end-users with a ranked list of potential qualifications; and c) present to end-users the above results in an easy and understandable way.

In the following, research activities that answer the needs of fostering students' mobility and that improve the quality of qualification offers will be investigated according to the three requisites identified above: a) knowledge base creation, b) integration of heterogeneous systems and definition of inference rules and c) visualization.

2.3.1 Knowledge base creation

In order to allow the automatic process of information, collected data should be expressed according to a shared formalism [Staab and Studer, 2004]. This task could be done at two different level of detail: a first one concerning the formalization of the way the information is expressed (that allows a machine to identify where to retrieve a given data), and a second level of detail allowing a software tool, once identified the searched information, to analyse its content (by describing concepts expressing its meaning).

The steps to be carried out for formalizing existing knowledge are well ex-pressed in [Tao et al. 2005]: in this work, the authors divided the knowledge lifecycle into 4 phases: knowledge acquisition (KA), that requires interviews with experts or desk analysis in order to develop a domain vocabulary of the most important concepts, knowledge management (KM), in which, starting from concepts identified in the previous phases, an ontology is built, knowledge annotation, in which resources from the domain are annotated with the ontological metadata, and, finally, knowledge reuse, achieved when new applications reuse the re-sources. In particular, in this work, for what regards the KA phase, learning do-main experts and teaching and learning materials were interviewed and investigated in order to identify key concepts, then, in the KM phase, Protégé[10], an ontology building tool, was used for the definition of concepts and hierarchical relationships among them. For the knowledge annotation phase, instances were generated and exported in OWL files, that were exploited in the knowledge reuse phase for recommending learning materials, or combining simple services in order to realize more complex customized functionalities.

While the above work provides an overview of the whole knowledge lifecycle process, and focuses mainly on the reuse of learning materials other authors investigated whether semantic tools could be used to support the personalization of training paths, with the aim of improving student's curricula. In particular, [Poyry et al. 2002] and [Poyry and Puustjarvi, 2003] shown how metadata could be exploited to support learners looking for higher education courses that match their needs: with this aim, the authors first investigated whether the Learning Objects Metadata (LOM)[11] could be used in order to describe courses provided by European universities, then suggested an extension of it, by identifying four aggregation levels (study material, study course, study package and study programme).

A more recent work, [Sampson and Fytros, 2008], realized that the LOM specification did not directly support the description of learning resources in terms of their relevance to competence-based learning programmes, and suggested an IEEE Competence Application Profile for tagging educational resources: the authors proposed to use Purpose (Nr 9.1) element to specify that acquiring a competence is the outcome of a learning object and the Difficulty (Nr 5.8) element to communicate the degree of difficulty of a learning object, and created a Competence category consisting of three main elements (title, description and context).

While the above works focus on methodologies for expressing the structure of information, other research activities aim at identifying best ways for expressing the content of a qualification in a machine-understandable format. This aspect is of primary importance in all the cases in which phrases syntactically heterogeneous, but with similar meaning, have to be compared, such as the identification of the best job applicant for a given job (since curricula are not necessarily expressed through

[10]http://protege.stanford.edu/

[11]http://ltsc.ieee.org/wg12/

the same terms that appear in a job offer), the comparison of different training offers, etc.

A first solution, mostly applied in the training domain, is presented by the Bloom taxonomy [Bloom, 1956]. In particular, the Bloom taxonomy was initially developed for educational assessment, and subsequently was exploited to evaluate courses according to the type of taught subjects. Bloom identified 6 levels of learning mastery (from less to more complex), namely knowledge, involving a mere recall of methods and processes, comprehension, representing the lowest level of understanding, application, requiring abstraction in particular and concrete situations, analysis, the identification of constituent elements of a communication and the understanding of relations between them, synthesis, the putting together elements so as to form a whole, and evaluation, the judgement about the value of methods for a given purpose. A revised version [Anderson et al. 2001], published in 2001, slightly modified the original taxonomy by inverting synthesis and evaluation concepts (renamed create and evaluate respectively), thus better reflecting engineering disciplines, and by expressing other concepts with a verb, instead of a noun. Moreover, the revised version defined also a new dimension, the knowledge dimension, describes as factual knowledge, conceptual knowledge, procedural knowledge and metacognitive knowledge. It is worth remarking that the this new representation allows training authorities to describe each element of a course as a pair verb-noun, thus defining 24 possible combination. An application of the Bloom's revised taxonomy is presented in [Spivey, 2007]: in this work the author presented how Digital Logic Design courses could be modeled according to [Anderson et al. 2001]. In particular, a matrix reporting the knowledge dimensions (nouns) and the cognitive process dimension (verbs) was drafted and, courses' objectives, activities, homework and quizzes were written in cells according to the their degree of difficulty and the kind of required knowledge. This representation allowed teachers not only to produce homework tailored on tough subjects, but made also possible to compare homework of common Digital Logic lessons, thus enabling, among others, the reuse of existing material.

Another example of how Bloom's taxonomy could be exploited for the comparison of qualifications is reported in [Bourque et al. 2003]: in this work the authors compared competences acquired by new graduates, graduates with four years of experience, and experienced software engineers by assigning them to one of the categories defined in [Bloom, 1956].

[Starr et al. 2003] presented a different nuance of the Bloom's taxonomy, since they introduced a meta-level structure: according to the authors, three meta-levels could be distinguished, memorization and basic understanding (beginner level), use or competent application (intermediate level), and design or creation and critique (expert level). Each one of them is represented by two phases, a *production of* *learning* and an *explanation* phase. Moreover, the authors identified an interesting phenomenon, namely *concept shifting*, that occurs when concepts describing a competence are inadvertently switched with a related, less abstract concept: in this case, the associated level of the Bloom's taxonomy may denote a different degree of difficulty, when applied to higher or lower-level concepts (e.g. the "iteration" concept at the Synthesis level implies the ability to design new ways to perform a loop, whereas the "for loop" at the same level entails only the ability to write loops to perform given tasks).

While the authors of the above works presented strategies based on the exploitation of the Bloom's taxonomy for a qualitative comparison of study courses, [Hoffman, 2008] defined a quantitative approach to compute the level of skills belonging to a qualification. In his work, the author started from the matrix reporting the knowledge dimension and the cognitive process dimension and, for each intersection, he defined a skill level explaining how much the given skill is acquired by students attending the course. In addition, he suggested to calculate also the center of gravity of both knowledge and cognitive process dimensions. This representation was extremely useful, since it quickly communicated the characteristics of a given course, as well as the level of the skills acquired by students.

The different approaches devoted to support learners, job seekers and job recruiters presented in this Subsection show how the exploitation of semantic tool could improve mobility and employability, as well as existing training offers. Devised solutions include a) the use of metadata for describing training courses [Poyry et al. 2002], [Poyry and Puustjarvi, 2003] and [Sampson and Fytros, 2008]; b) the exploitation of the Bloom taxonomy for the description/comparison of learning paths, so as to allow students to improve their employability by personalizing their curriculum [Spivey, 2007], [Bourque et al. 2003], [Starr et al. 2003] and [Hoffman, 2008];

The utilization of metadata is particularly useful in those cases in which information is unstructured (such as courses from different providers, or curricula ex-pressed with a variety of structures): in this view, they could improve search and comparison phases, by providing more correct results. However, if only metadata are used, users can only perform keyword-based searches, since the system could not be able to compare sentences expressed through different terms. In order to perform such a work, taxonomies or ontologies defining relations among terms should be exploited. In this view, the Bloom taxonomy is certainly useful for the description of skills, since it identifies families of verbs grouped according to a degree of complexity. However, it lacks of relations among terms belonging to the same group, and it is still limited to a small set of concepts. A possible way to overcoming this limitation could be the extension of it in a given context, by the creation of an ad-hoc taxonomy/ontology for the domain of interest: this solution could allow a software system to automatically deal with job offers, qualifications, etc. expressed in heterogeneous way. However, when a wide context like qualification comparison is considered, this solution is no more sustainable, since an ad-hoc taxonomy/ontology should be developed for each working/learning domain, thus resulting an extremely time-consuming activity. Research and future works in this field should then focus on the adaptation of existing ontologies/semantic thesauri, such as WordNet, DBpedia[12] [Auer and Lehmann, 2007], etc. and to their exploitation for automatically annotate curricula and training courses. The only drawback of this solution is that, in some contexts, the above ontologies are not sufficient, since they lack of concepts or relations. Hence, applications allowing users to eventually extend them should be developed. Moreover, since usually a word could assume one or more meaning, according to the context in which it is used, the possibility to automatically annotate the inserted data according to information already provided by users (e.g. the domain of interest, etc.) should be investigated

2.3.2 Integration of heterogeneous systems and definition of inference rules

While Section 2.3.1 presents different strategies for the creation of a knowledge base collecting curricula, different approaches could be devoted to the integration of existing academic management systems and to the definition of rules for matchmaking (a process that queries a knowledge base and returns all the elements that potentially match the requirements expressed by the user), in order to support students looking for courses to attend for improving their curriculum, or training institutes willing to improve the training offer. As presented in [Ronchetti and Sant, 2007], ontologies could be used for managing, inspecting and monitoring a full study courses, since they could allow a system to verify overlaps between courses, to find out areas which are not covered and to analyse possible synergies with courses offered in other schools. The authors started from the ontology extracted by [Saini and Ronchetti, 2003], based on the ACM Computing Curricula 2001 for CS (CC2001) [IEEE and ACM, 2001], a comprehensive work defining Computer Science curricula for undergraduate students, specifying, among others, prerequisites and syllabus for courses. In the CC2001 view, each course could be defined by topics (the smallest-grain elements, divided into core and elective), units (collections of topics) and areas (collections of units). The objective of [Ronchetti and Sant, 2007] was to analyse syllabi of the courses provided by the Computer Science Bachelor at the University of Trento, Italy. The pursued approach consisted in matching all syllabi against the ontology, asking teachers to identify taught topics, and showing the result of the mapping. Results of these activities allowed trainers to identify not covered areas,

[12]http://dbpedia.org/About

thus improving the training offer. A different methodology, developed in order to enable learners to integrate classes from other institutions into their curriculum, in the Bologna Process view [Bologna Declaration, 2001], is shown in [Hackelbusch, 2006]. In this work, the author presented a system providing students with a ranked list of classes offered by other academic institutes, by including only classes that were identified as interchangeable, from the organizational and the semantic point of view. With this aim, a curricula mapping ontology, used as a common basis for formalizing academic programs, was developed. However, in order to find similar classes, a mere comparison on the title of modules was not sufficient, since modules with the same title could contain completely different subject matters. Hence the author proposed to exploit methods for indexing texts in order to summarize modules content.

While the above work starts from the definition of an ontology and requires a formalization of academic programs, the strategy proposed in [Cubillos et al. 2006], [Cubillos et al. 2007a] and [Cubillos et al. 2007b] starts from the assumption that, due to political reasons, in some cases, the creation of a common ontology could not be possible, hence, local meta-ontologies, allowing an automatic tool to compare heterogeneous qualification systems, should be developed. The aim of these works was the definition of a methodology exploiting meta-ontologies to tackle the problem of integration of qualification systems in the Vocational Education and Training (VET) context, in order to improve the transparency and mobility of students across Europe. In particular, an initial document gathering phase was carried out to collect and analyse all the relevant information, then, a UML model[13] was drafted. Finally, a third step consisting in the construction of templates and in the compilation of case studies, was performed. Based on the above documents, an ontology built with OWL language were developed.

In particular, for what it concerns the documents gathering phase, information about the national educational system (reporting an overview of the general educational system of the country), the post-secondary non-academic education (detailing different types of school providing this kind of qualification), the post-secondary non-academic education system analysed for the project (a deeper analysis for selected systems) and the profiles constitutive parts and description of concepts (a detailed study on how a profile was articulated in terms of competences, modules, etc.), were collected. The UML formalization phase was carried out in order to provide a formal and easy to read representation of gathered documents: in this phase classes and attributes for each qualification model were detected, and relations among them and foreign qualifications were drafted. At the end of this step, one UML diagram for each couple of Countries to be compared, was provided. However, since this

[13] http://www.uml.org

representation was not familiar to all partners involved in the development of the tool, formalized information was also structured in MS-Word templates, in order to be checked also by a non-skilled person. These templates were also used for the compilation of study cases.

Once meta-ontologies were defined, qualification profiles were annotated by referring them to a common glossary of concepts, and, for each concept, a weight expressing its relevance was inserted. [Gatteschi et al. 2009] aimed at exploiting the above models in order to help students to identify, starting from their missing competences, courses to attend in a foreign institution, and proposed two different strategies for matchmaking. In particular, a fist approach, exploiting Country-to-Country relations was pursued. However, it was pointed out that this strategy is not suitable when the number of qualification systems to be compared is high, since, for each comparison, a meta-ontology has to be developed. Hence, a Country-to-EQF strategy has been suggested: according to this approach, constituting classes of a qualification have to be referenced to classed identified by the EQF (such as learning outcome, knowledge, skills, competence, etc.), thus exploiting the EQF as a translator device.

As shown by the works presented in this Section, tools developed within the Semantic Web initiative could be used for performing the following activities: a) analysing and verifying training offers [Ronchetti and Sant, 2007] and [Saini and Ronchetti, 2003]; b) comparing training modules in order to allow learners to personalize their curricula [Hackelbusch, 2006]; c) express heterogeneous qualifications through a common formalism [Cubillos et al. 2006], [Cubillos et al. 2007a], [Cubillos et al. 2007b] and [Gatteschi et al. 2009]. For what it concerns the usage of ontologies for verifying eventually missing competences, in a training module, the work presented in [Ronchetti and Sant, 2007] turned out to provide interesting results. However, an approach like the one shown by the authors requires an ad-hoc ontology for each training domain, hence the feasibility of this strategy in different sectors should be investigated. The system presented by [Hackelbusch, 2006] is able to compare modules organizational requisites and tries to investigate whether a comparison could also been done on the content of modules themselves. However, this approach seems to be able to work only on courses owning the same structure. In this view, learners could benefit from the exploitation of a meta-ontology to link heterogeneous training systems, through the identification of a common framework [Gatteschi et al. 2009]. In fact, even if this solution requires the definition of shared framework, as well as a considerable amount of work from experts of the training domain of different countries in order to create ontologies for each couple country-common framework, it could be used as a translator device when learners want to compare foreign qualifications for finding interesting training paths. The definition of rules for matchmaking could simplify the above activities, that could be carried out automatically, instead of on a manual basis.

2.3.3 Visualization

While Section 2.3.1 and Section 2.3.2 focus on the representation of knowledge in a machine-understandable way and on strategies to perform automatic reasoning on it, another aspect to be considered is how huge amounts of data, such as the ones usually characterizing a context like the one analysed in this Chapter, could be effectively presented to end-users.

In fact, even though a system for the comparison of qualifications could provide the users with a list of ranked results, they could be interested in analysing why a given qualification obtained the specific ranking position. Consequently, in the following, different approaches to representation of qualifications and curricula will be investigated.

Several research activities have been carried out in order to provide a graphical representation of qualifications outcomes or relationships between courses. A first approach is presented in [Gestwicki, 2008]: in this work, the author suggested a curriculum visualization application in which a curriculum was modeled as a directed graph where courses were represented by nodes and relationships between them were depicted by edges. By looking at the graph, it was then possible, for students and program administrators, to identify the flow of a curriculum in terms of courses prerequisites, or amount of time needed to satisfy courses requirements. According to this representation, elective courses were represented by shaded nodes, whereas required courses were depicted as unfilled nodes. Nodes were then linked by different types of arrows, representing, as a matter of example, prerequisites. By selecting a shaded node, and by analysing arrows pointing to it, it was possible to identify prerequisites, and, by finding the longest path between nodes, the user could determine the number of semesters the fullfilment of courses requirements would take. Another interesting strategy, exploited to manage and maintain Medicine curricula, is reported in [Dexter and Davies, 2009]: this research aimed at easing the continual process of review characterizing curricula requiring frequent revisions, and adopting the Problem Based Learning approach. In particular, the authors, in a first stage built a visual model of the curriculum by exploiting the UML formalism, thus identifying, among others, the competency class, then tagged each competency by assigning it values for different dimensions, and built the model as an ontology in OWL. Finally, in order to visualize the created knowledge base, they suggested a visual metaphor derived from the London Underground map: on this map, elements of the ontology were represented as stations, whereas object properties (i.e. relations) were depicted by lines. Similarly to the London Underground map, not all stations are linked to all others, and there could be different routes that may be taken between stations, hence the user can browse the knowledge base by selecting a station and by following trails connecting related items. While the above works adopt a bi-dimensional representation of qualifications, other authors suggested a three-dimensions depiction. In fact, [Hoffman, 2008] represented skills composing a learning objective as blocks defined by a triple of values characterized by the knowledge dimension (x axis), the cognitive process dimension (y axis) and the skills level (z axis). According to this description, the height of a block communicates the degree of acquisition of a given skill, whereas the spatial position gives information on its level of complexity. [Sommaruga and Catenazzi, 2007] took a step forward in this direction, by representing university undergraduate education programmes in a 3D environment, based on a city metaphor. In this view, each department is depicted as a district, whereas each curriculum offered by the department is represented by a block, divided in 6 areas, one for each semester (since bachelor courses are three years long). Each block contains one or more buildings (modules of a curriculum), whose height and width are proportional to the number of credits and to the duration of the module, respectively. An ad-hoc exploitation of colors (different colors for each department, and different nuances for each semesters) completed this representation, by making easier to distinguish main characteristics of a training path, and by allowing users to quickly compare different education programmes. While authors of the above works suggested strategies to visualize relations among courses and prerequisites, and to compare training paths by analysing graphical depiction of skills or modules, an interesting field of research is represented by the visualization of qualification content. In fact, in order to compare qualifications, also taught concepts should be took into account. A quick way to produce a graphical representation of main subjects provided by a training institute is to create a tag cloud (a visual overview of textual data, often corresponding to a set of tags in which the font size used for drawing the tag is generally linked to its frequency) based on courses specifications.

As it has been shown, there are several strategies to quickly depict main aspects of a qualification/curriculum vitae, and they could be grouped in the following groups: a) approaches exploiting a graph-based representation of information through nodes and relations among them [Gestwicki, 2008] and [Dexter and Davies, 2009]; b) three-dimensional representations of skills provided by a qualification [Hoffman, 2008] or of the composition (in terms of modules, credits, etc.) of education programmes [Sommaruga and Catenazzi, 2007]. Since the visualization of information strictly depends on the type of data to be displayed and on the characteristics of end-users, it is difficult to identify a priori the best approach to be pursued. In general graph-based representations turned out to be able to provide an overview of a qualification, but there is the possibility that some end-users would not be able to fully understand them. However, the approach suggested by [Dexter and Davies, 2009], consisting in the creation of a visual metaphor of the London Underground map, could allow even non-skilled users to comprehend them. Similarly, three-dimensional representations have the advantage to communicate to the user different information about a given qualification/curriculum, but they should be simplified in order to be fully understandable by each involved actor, like in the work of [Sommaruga and Catenazzi, 2007].

2.4 Semantics and job matchmaking

The Internet represents today the backbone that conveys the huge information flow generated by modern job seeking and job recruiting tasks [Ployhart, 2006]. In fact, many companies have endowed their corporate portals with work for us sections, and a number of web tools have been developed to help job seekers organize their résumés in a predefined way possibly easing the process of eliciting their attitudes and abilities. Moreover, many dedicated job portals have appeared, allowing employers to post announces about open positions and supporting job seekers in the process of matching their expertise with such offers [Bizer et al. 2005]. Even social hubs originally developed for letting users strengthen their professional networks and share generic personal content lately added the possibility for subscribers to describe themselves by means of a set of skills and match their profile with a number of job opportunities. Recent studies showed how the web is quickly becoming a virtual marketplace where an ever-growing number of job placement procedures are initiated [Buckley et al. 2004] [Bourse et al. 2002] and global portals count so many curriculum vitae and job postings recorded in their databases that e-recruitment is considered by now as one of most widely practiced e-business areas in organizations [Lee, 2011] [Lv and Zhu, 2006] [Monster, 2012].

Like in almost any other domain, the enormous communication opportunities offered by the web contributed at reducing information asymmetries between actors involved in the job seeking and job recruiting scenarios, which started using the Internet as the main channel for exchanging information. While the number of users commenced to grow, the amount and complexity of data to be considered started growing as well, and it became practically unfeasible for recruiters and job seekers to manually sort out data regarding relevant candidates or positions, which are today scattered on a number of separate systems and expressed in many different ways (as structured, semi-structured and unstructured résumés, job postings on corporate websites, LinkedIn and Facebook profiles, etc.).

Therefore, various approaches for automatic information processing necessarily started to be borrowed from other specific application fields and experimented for dealing with job seeker and employer-related data. This way, the concept of matchmaking, which generally refers to the act of aligning the offer and demand of some kinds of goods in (virtual) marketplaces started to be exploited in the context of e-job seeking and e-recruitment. Here, parties involved are trying to sell and buy delicate types of goods, i.e., skills, know how, attitudes, etc., in a open and distributed market, that is the labour one. Though, as it will be shown in the following, almost any solution considered was capable of addressing somehow the challenges introduced above [Lv and Zhu, 2006] [Fazel-Zarandi and Fox, 2012] [Colucci et al. 2007], the use of semantics really proved to represent a holistic way for dealing with the overall picture [Mochol et al. 2007a] [Colucci et al. 2003]. The superior advantages of techniques based on this latter approach are witnessed today by the number of works reported in the academic literature and, most significantly, by the growing number of commercial job portals advertising their semantic-aware search and ranking features.

It is impressive to realize what is the effort that has been put not only by the researchers' community, but also by organizations in general, onto the application of semantic solutions to the labour market world. This is due to the fact that the tasks performed and outcomes expected by a job seeker or by a recruiter working on a job matchmaking tool are exactly aligned with those of one the semantic-based knowledge management systems above like, for instance, a semantic search engine.

In fact, when a job seeker drafts a résumé he or she has the goal of making all the acquirements gained in any study or prior work experience visible to possibly interested parties. Similarly, human resources managers want to rapidly and effectively sort out received applications to elicit those characteristics that, based on the requirements set, make a given individual the best candidate for the position to be covered.

Unfortunately, though the two sides of such matchmaking deal are talking about the same subject, they often do not talk exactly the same language. For instance, a job seeker might describe him or herself by making reference to a professional occupation, like *Java programmer*, whereas another individual might mention his or her personal abilities, e.g., in the form of *being able to write a program in the C++ language*. Similarly, a company could simply post the title of the open job position, like *software developer*, or specifically detail attributes required, like *programming with OOP languages*.

A human resources manager expert of the field would immediately realize that Java and C++ are both OOP programming languages, as well as that if someone is able to write a computer program using a kind of programming language can be considered a software developer. That is, for solving the matchmaking problem and finding the best candidate, he or she would implicitly and natively exploit information about the conceptual and contextual meaning of and relations between terms concerned, which in general could make reference to information expressed, among others, with different words, at different levels of details, with heterogeneous degree of completeness, etc.

As said, the goal of semantics and semantic processing is exactly in the same direction, i.e., to empower machines with the ability to automatically extract and analyse relations between terms and concepts to intuit and infer new knowledge in a way that should be as close as possible to typical human reasoning. In a semantic tool, concepts like programmer, program, programming and software developer as well as OOP, C++ and Java languages would have to be first identified. Then, links among concepts would need to be explored, which could be from rather trivial to rather complex. As a matter of example, concepts like programmer, program and programming would be linked by straight lexical relations. Concepts like OOP, C++ and Java could be framed in a possibly incomplete hierarchical structure, with implications between concepts (the Java and C++ languages belong to the family of OOP languages and represent a specialization of the more general category, though they do not exhaust it). Furthermore, they would be also linked to the act of programming, i.e., of writing a program and, quite intuitively, to the role of programmer, who is generally in charge of writing programs. This way, the relation between the above elements and a software developer job position would be elicited.

It could be worth adding that, when the job matchmaking scenario is regarded from an Internet-wide perspective, in most of the cases job seekers and employers are not even talking in the same place. In fact, the chasing evolution of web technologies and the rush to the adoption of the most recent research achievements paradoxically resulted in the creation of a number of isolated and non-interoperable e-job seeking and e-recruitment islands [Bizer et al. 2005], characterized by platforms based on heterogeneous interfaces and approaches. In this sense, the use of semantic technologies as a way for formalizing data exchange could be seen again as one of the key ways for achieving the necessary integration in the e-human resources management domain.

In the following, the application of semantic-based solutions to job seeking and job recruiting domains will be analysed in details. In particular, possible applications of semantics to the description of job profiles and curriculum vitae, to the elicitation of required or owned competence, skill and knowledge elements and to the implementation of automatic solutions for computing the job offer and job demand matching will be presented. Finally, the discussion about advantages demonstrated by or expected from semantic tools in the job matchmaking context and in other envisaged application scenarios will be balanced with an analysis of limitations affecting current techniques and implementations as well as of evolutions planned.

2.4.1 Semantics in the job seeking and job recruiting scenarios: an overview

The idea of exploiting the concepts of semantics to deal with job seeking and/or job recruiting issues is rather old, already. In fact, the first solution for matching job offers and job search requests dates back to 1990, when the author of [Vega, 1990] presented a tool for enabling professionals to quickly and efficiently search for job positions advertised in the online database of a French newspaper. Candidates were presented with a user interface for entering their curriculum vitae using natural language statements in a structured template. A matching machine was responsible for extracting normalized terms from advertised job offers and collected curriculum vitae, which were modelled using object analogy. Starting from the initially identified terms, semantic-pragmatic rules were used to spread the initial knowledge and create a new set of terms, which were processed by means of intersection and classification rules. Indeed, at that time user interaction was rather simplified, since the access was based on text terminals that were quite far from today rich web clients. Moreover, the focus was only on job seekers, and the system was only meant to support users in obtaining an ordered list of job positions maximizing the match with the curriculum vitae being entered. Lastly, the technologies now supporting the Semantic Web initiative were still far to come, and solutions implemented worldwide to standardize professional activities and abilities for job occupations, learning outcomes mobilized in formal learning paths and involved in e-learning scenarios, among others, were still to be defined.

Nonetheless, such embryonal work was at least mentioning already most of the challenges involved in today's semantic matchmaking in the labour context, namely the dual jobseeker – job recruiter perspective and the importance of smart tools for accompanying data entry in both the domains, the need to represent user-provided information by linking terms used to categories of formalized concepts with various kinds of relations between them and, lastly, the role of semantic rules for reasoning and resolving ambiguity, incompleteness, heterogeneity, etc. into knowledge collected (Figure 2.5).

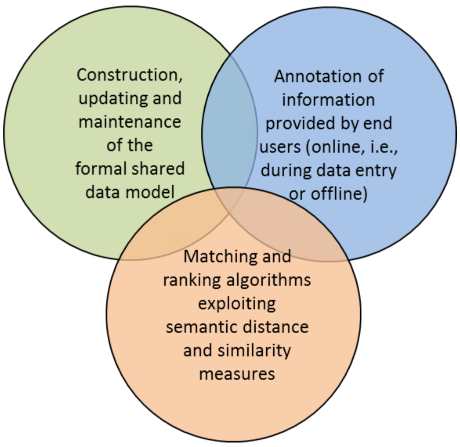

Figure 2.5. Challenges of today's semantic matchmaking in the labour context

The above focuses could be confirmed by a review of the academic literature produced as well as of the commercial products released in the last two decades and, especially in the last about 10 years that followed the definition of the concept of Semantic Web. As it happens in almost all the domains, in many cases research ideas and prototypes originated in the academic sphere have been transformed into business opportunities and public services. However, though on the one side research works almost equally considered both the job seeker and employer perspectives, commercial platforms mostly focused on the latter actors because of the enhanced business opportunities. Thus, as a matter of example, various online recruitment systems sell résumés search service to companies, which can post a job offer and disclose details of suitable candidates by paying a certain fee. Even social platforms not explicitly born to support these kinds of functionalities started selling APIs for performing recruitment-oriented queries over subscribers' public data.

Independent of the perspective, a huge number of works concentrated on how to formalize the knowledge embedded into curriculum vitae and job offers collected. Various kind of domain ontologies were therefore created, starting from existing thesauri and taxonomies and later involving a number of domain experts who are generally asked to perform a significant amount of manual activities aimed at designing, merging, updating and maintaining the required knowledge bases.

Then, many efforts were spent to study how to actually implement the matchmaking process. As said, to fully benefit from the automatic processing capabilities of semantic technologies, information regarding end-users, either job seekers or employers, need to be made explicit, i.e., linked to the structured data defined. This is the only way to overcome ambiguity, generality and incompleteness, etc. associated with the interpretation of users' words and sentences. The various solutions developed often differentiate based on the strategy pursued for implementing this step. In fact, in some cases semantics is exploited to support data insertion, by acting as a facilitating means for driving the résumés or job offers compilation. In other cases, semantic relations are exploited only at a later processing stage, i.e., to widen the domain space explored in matching job offers with demands. Finally, there are few situations where the considered technologies play a dual role, by being exploited as a way to let the users formulate and/or adjust their queries, which that are then adopted to find the best match in a search space that is explored in a semantic-aware way. The overall picture of a possible comprehensive framework supporting semantic job matchmaking as it has been theorized or as it would result from the integration of works and products available is illustrated in Figure 2.6.

The main advantages identified in such a framework as perceived by end-users are related to a shortening of employment and hiring times and a reduction of costs for both recruiters and job applicants. In fact, with respect to traditional job seeking and job recruiting scenarios, with the devised approach time and money can be saved at any stage, from job advertising, to applications processing, candidates evaluation and screening, as well as from job search, to application submission and interviews execution. By the way, a larger market can be reached by both résumés as well as job

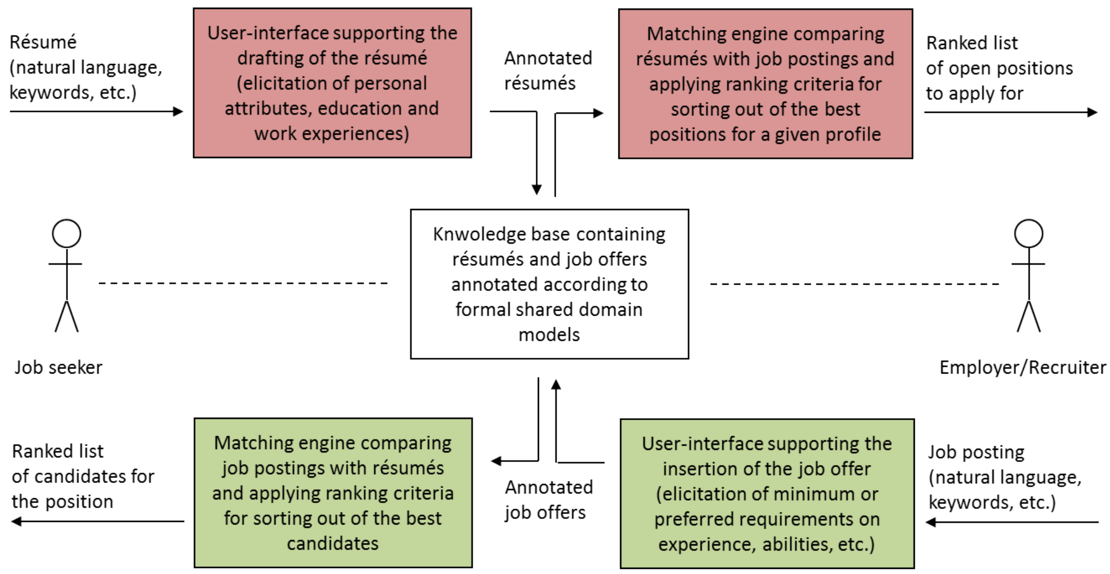

Figure 2.6. Framework supporting job matchmaking

postings. The publishing of job offers online is even regarded by job seekers as a sign of good health for a company [Buckley et al. 2004] [Lee, 2011]. It is worth saying that, besides serving as a technological means for an effective paperless processing of a huge amount of data in a short time by providing indications about the possible best matches according to certain criteria, technologies discussed had the additional advantages of relaxing the search abilities and the knowledge of a specific domain being requested to end-users by both traditional or alternative techniques. That is, the job seeker and the recruiter are no more requested to find the right keywords or to be experts of a particular industry or skill set. This way, further opportunities for finding a job or acquiring a human resource in a ever more inclusive society are actually opened.

In the following, the various aspects mentioned above about knowledge management are specifically considered, by tackling aspects related to representation, annotation and, lastly, exploitation of job seeker and employer-related data. In this respect, the most relevant works about semantic matchmaking reported in the academic literature are first considered. Then, the functionalities of publicly available products are analysed and compared with the dominant approaches just identified.

Knowledge representation in the field of job matchmaking

Almost all the market products and research projects reported in the literature have stressed the importance of a common and formal language for dealing with the subjects involved in the job seeking and job recruiting processes. Indeed, such a language has to be shared between the two sides concerned, i.e., job seekers and employers. That is, it has to be capable of supporting the description of the elements of interest for both parties, i.e., curriculum vitae on the one side, and job offers on the other. This need requires to first of all set out a standard for collecting data. In this perspective, it is worth observing that, though the way a job offer is structured could vary based on the country, sector and company considered, many postings make reference in principle to established job positions with known requirements, which nonetheless can be further detailed by specific corporate needs. Similarly, the way a résumé is drafted, at least in terms of macro-fields, is in many cases quite structured. This is also due to efforts carried out even at the institutional level to establish a common set of information generally expected. As a matter of example, in the tool named European Curriculum Vitae defined by the European Commission in the framework of the Europass portfolio[14], several broad sections are identified, such as personal data, work experience, education and training, languages, and additional information, such as publications, personal interests, etc.

But, most importantly, what the institutional tools, the research works in the literature and the e-recruitment platforms already online agreed upon is the key role played by the concept of competency in the description of both employer's and job seeker's documents.

If fact, the competency logic is now quite consolidated in the field of human resources management, since competency is regarded as the real capital of a company, especially when knowledge intensive tasks such as decision making, strategic planning, creative design and the like have to be dealt with [Taubner and Brössler, 2000] [Younker, 1998] [Legge, 2005]. Thus, the advent of the Semantic Web has been considered as an opportunity for developing this logic also in the field of e-job seeking and e-recruitment, by carrying out the important formalization processes required. As a matter of fact, a common trait of almost all the strategies devised to deal with the description of job-related acquirements (in the case of a job seeker's résumé) or requirements (in the case of a job offer) is represented by the attention paid to precisely define competency achieved or to be mobilized, by also making reference to the other outcomes and elements of formal, non formal and informal learning paths. Hence, in most of the works considered, job seekers' profiles and job offers are regarded as "containers" of competency items to be reciprocally matched.

Despite such a consensus of the role of competency, the concept itself assumed many different definitions, depending on the context the semantic approach was developed into and the actors involved (employment offices, trade unions, chambers of commerce, recruitment companies, education and training institutions, sectoral organizations, standardization bodies, etc.).

[14]http://europass.cedefop.europa.eu/en/documents/curriculum-vitae

For instance, in [Yahiaoui et al. 2006] competency is referred to as *a set of knowledge used to accomplish a task*. According to the authors, it can show up as an aptitude (behaviour) or can appear as a scientific/technical ability (know-what or know-how), which in turn can be specific or general (depending on the domain addressed). In the authors' approach, which chose as a target the ICT domain, a competency object (e.g., a specific topic or a software artefact) is associated a competency level: basic, application, master ship or expert. Finally, aptitudes are described by making reference to the CIGREF[15] classification, which has been developed by a consortium of 130 French companies committed to support the diffusion of information systems and technologies in the labour market and the society in general. The model used to formalize competency is quite simple, since the goal is to maintain the complexity associated with the management of descriptions of job resources low and to achieve a good trade-off with efficiency. A similar approach is adopted in [Lv and Zhu, 2006], though in this case the term used to refer to achievements or requirements is skill. Skills play a key role also in [Fazel-Zarandi and Fox, 2009], where competencies are modelled as skills with a proficiency level associated, where the level is determined by a knowledge degree and a number of years of work experience. Work experience in turn is modelled by making reference to the position in the organization, the duration of the job and competencies associated. In [Fazel-Zarandi and Fox, 2009], learning activities are also described, by making reference to learning paths with associated competencies as outcomes. It is worth observing that, although in the above examples a distinction is actually made between competencies, skills and related terms, in many cases they are used as synonyms, and intentionally confused also with competences, abilities, know-how, etc.

A slightly more sophisticated approach is considered in [De Coi et al. 2007], where the definition of competency in [Cheetham and Chivers, 2005] is used, that is *effective performance within a domain/context at different levels of proficiency*. In this case, the confusion in the literature is generally related to the interpretation of concepts like proficiency, level and context, which represent the three perspectives that have been considered by the authors in their knowledge modelling stage (for instance, "Fluent business English" would represent the competency "English", the proficiency level "fluent" and the context "business").

Another definition of competency is the one given by HR Consortium[16], an independent organization aimed at promoting the use of XML-based solutions (known as the HR-XML suite) in the field of human resources management. In this context, competency is defined as *a specific, identifiable, definable, and measurable*

[15] http://www.cigref.fr/

[16] http://www.hr-xml.org/

knowledge, skill, ability and/or other deployment-related characteristic (e.g., attitude, behaviour, physical ability) which a human resource may possess and which is necessary for, or material to, the performance of an activity within a specific business context. This definition is adopted in a significant number of works like [Bizer et al. 2005] [Fazel-Zarandi and Fox, 2010] [Mochol et al. 2004], among others. In [Bourse et al. 2002], the authors exploits a competency model that is very similar to the HR-XML related one, where a competency corresponds to *a set of resources (i.e., knowledge and/or behaviour and/or more basic competencies) that is mobilised in a particular context for reaching an objective or fulfilling a mission.* Also in this case, like in a number of other works, three fundamental characteristics emerge from the definition, i.e., resources, context and objective. The authors of [Bourse et al. 2002] consider three types of resource categories, i.e., knowledge (related to the education domain), know-how (related to personal experience and working conditions, synonym of skills, operational capacities or experiments) and behaviour (individual characters, including human traits, qualities and attitudes). Based on the definition, they consider the context, related to the environment in which the competency is situated, and to the objective. In the authors' vision, competencies are described by means of a quintuplet making reference to the elements identified in Figure 2.7.

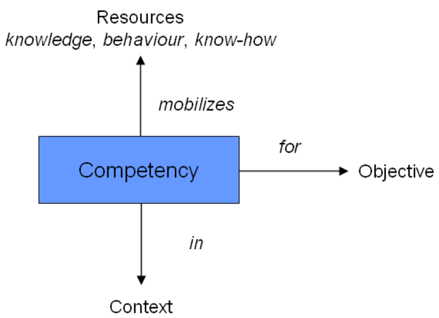

Figure 2.7. Description of a competency

In [Gatteschi et al. 2010], [Gatteschi et al. 2011] and [Pernici et al. 2006] the more general concept of learning outcome is exploited, defined in the EQF as *a statement of what a learner knows, understands and is able to do on completion of a learning process.* Learning outcomes are specified in terms of knowledge, skills and competences. Knowledge can be theoretical or factual, and encompasses body of facts, principles, theories and practices that are related to a field of work or study. Skill means the ability to apply knowledge and use know-how to complete tasks and solve problems, and can be cognitive or practical. Finally, competence means the proven ability to use knowledge, skills and personal, social and/or methodological abilities, in work or study situations and in professional and personal development and it is considered in terms of autonomy and responsibility. Moving from the above definitions, as well as from the guidelines provided in Europass for the description of competences[17], in such works a knowledge is regarded as a set of Knowledge Objects (KO), whereas a skill is represented as a KO *put into action*, i.e., as one or more pairs KO – Action Verb (AV). Finally, a competence is identified by means of a triple KO – AV – CX, which describes the ability of putting into action a KO *in a specific context* (CX). The overall formalization approach can be represented as in the diagram below. It is worth observing that, in Europass, a glossary of action verbs to be used in describing competences is also provided.

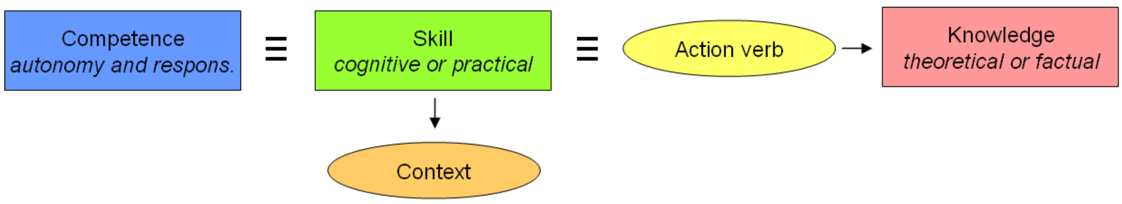

Figure 2.8. Relation between knowledge, skill and competence elements

Based on one of the definitions above, in the various experiences analysed specific taxonomies and/or ontologies tailored to the job matchmaking domain are generally defined, supporting a formal description of information contained in job seekers' profiles and job postings. The domain knowledge is often "created" by merging and updating existing formalizations and differences are mainly in the complexity of the underlying competency model, in the number of job seeker and job offer-related data considered and in the depth of details used in the modelling.

It is worth observing that in almost all the works, this latter step is considered as the most critical one, because of its importance as well as of the significant human effort required. For instance, in [Reich et al. 2002], the process of developing the domain ontology is presented as an iterative, incremental and evaluative process based on brainstorming and mind mapping, where domain experts (often referred to as *knowledge engineers*) are initially provided with just a simple top level ontology and some constraints on the number of branches and on the depth of the model.

In [Lv and Zhu, 2006], the authors focus on the definition of a skill ontology model for a particular sector of interest. The model is basically a graph made up of skill nodes and is-a or part-of edges (e.g., a C++ experts is quite familiar with C, whereas a C expert knows something about C++ but does not master it), with directions between nodes. Each node depicts a particular skill, whereas nodes define

[17]http://europass.cedefop.europa.eu/en/documents/european-skills-passport/certificate-supplement

how two skills are correlated. Edges are assigned a weight, describing how a part-of relation contributes to the meaning of the related node. In [Fazel-Zarandi and Fox, 2009], a skill ontology is used to describe a job seeker as "equivalent" to a set of skill statements, endowed with suitable proofs (degrees and/or previous work periods). Similarly, job offer requirements are modelled as constraints or desired (nice-to-have) skills or degrees. The authors model the skill ontology as in [Lv and Zhu, 2006], by using is-a and part-of links as well as the alternative-for symmetrical relation. In [Yahiaoui et al. 2006] a so called electronic recruitment ontology is developed by linking sub-ontologies according to the meta ontology-based approach defined in [Fernandez et al. 1997], by describing persons (and personal attributes), diplomas (and related training), and the jobs and by eliciting the connections with the underlying competency elements (represented by ICT-related topics organized in a hierarchical structure, each linked to a weight). It is worth saying that the creation of hierarchies of concepts based on object analogy as well as semantic and pragmatic relations had been exploited in [Vega, 1990] already.

In [Bourse et al. 2002], curriculum vitae and job offers are considered as a synthetic view (expressed in natural language in terms of qualifications, work experiences and extracurricular activities) of a richer network of competences. Thus, together with a rich competency model, a number of other semantic formalizations are exploited, encompassing sector ontologies (developed by the actors of a specific sector and its associated professions), enterprise ontologies (that could include, for a given profession, the description of additional tasks that are performed within a particular company) and behaviour ontologies (dealing with the schematization of individual traits).

In [Bizer et al. 2005] the authors analyse how the shared vocabulary to be used by job seekers and employers could be derived from standards already in use in the recruitment domain for describing occupations, required skills and educational background. Instead of defining new models from scratch, they consider the wide range of standards already defined and study how to integrate them into a so called (quasi comprehensive) human resources ontology.

A similar approach is exploited by the same authors in [Mochol et al. 2007a]. The starting point is represented by the HR-XML suite that, as said, is exploited also in a number of other scenarios (possibly with some variations). The HR-XML data format consists of tens of schemas dealing with the description of information pertaining a number of areas and business processes. Based on HR-XML, the authors study a methodology for combining it with a number of taxonomic models designed for classifying occupations, sectors, professional activities, etc. In particular, they take into account the Standard Occupational Classification System (SOC)[18], the

[18] http://www.bls.gov/SOC/

North American Industry Classification System (NAICS)[19] and two other models specific for the national context. Finally, they describe skills by making reference to an extended version of the ontology defined by the University of Duisburg-Essen[20], by also developing an ad hoc model for detailing personal information. They do not consider education-oriented classifications. In [Mochol et al. 2004], the authors start from HR-XML and (re)define sub-ontologies for dealing the various thematic clusters of interest, regarding industry sector of the job position, job position details, etc. HR-XML is the starting point also for [Fazel-Zarandi and Fox, 2010]. Here, the authors rely upon the definition of competence and its relation with skill, knowledge and activity performance by the HR Consortium. In particular, they model in the ontology the latter relation by formalizing the fact that a skill (synonym of competence, in this case) *enables* an activity. Then, they define specific tools and technologies as resources *required-by* an activity. They additionally identify *related-to* links between skills, thus indicating that skills are related to each other in the domain considered. In the same work, the authors also define the concepts of learning activity and of work experience and use the *has-outcome* and *has-experience* relations to elicit the link between formal learning and working period with acquired or expected skills (with a certain level associated). Finally, they model the concept of job posting, by defining, among other, the has-requirement relation to make the needs explicit in terms e.g. of skills, knowledge level in the use of a given tool or technology. The approach in [Ionescu et al. 2012] also moves from HR-XML and then deals with the issue of creating an ontology specifically tailored to the ICT field and to the database area, in particular. The authors specifically study how to populate the ontology with concepts and relations. To make an example, the database ontology is divided into theory foundations and technical aspects. Below, two new sub-ontologies are created, named data modelling and database management. Each subdivision is assigned a score indicating the strength and contribution of that sub-ontology to the upper concept.

In [Gatteschi et al. 2010] and [Gatteschi et al. 2011], moving from the definition of competence as a skill mobilizing certain knowledge in a specific context, the authors model knowledge items, action verbs and context elements as three separate taxonomies. Then, competences are formalized by establishing relations between the three hierarchical structures, thus creating an EQF learning outcome-aware ontology that is then exploited for performing the required annotation. In [De Coi et al. 2007], the authors carry out an analysis on the various approaches experimented for modelling relations between competences. Based on such analysis, they observe that in many works dependencies/equivalences are modelled, including the composition

[19]http://www.census.gov/eos/www/naics/

[20]http://duepublico.uni-duisburg-essen.de/servlets/DozBibEntryServlet?mode=show&id=3800

of complex competences from simple ones. However, they realize that combination and weighting of competences is not clearly defined. That is, they state that relations between competences should not be modelled, especially if the proficiency level and the context are not considered. To make an example, the knowledge of statistics could be a requisite for becoming a computer scientists or a sociologist, However, the proficiency level required and the competence would be extremely different. Thus, they model the three perspectives separately. Competencies are modelled without relations. Subsumption relations are in turn identified between proficiency levels. Finally, context elements are modelled by means of a tree structure. Competences, consisting of the above elements, can be composed to create aggregate competences.

In [Gómez-Pérez et al. 2007], the problem of résumés and job offers description is addressed from the point of view of employment services interoperability with the aim of supporting the creation of a truly shared job marketplace. Like in the previous experiences considered, the authors move from the consideration that achieving an overall consensus with a comprehensive ontology created from scratch would be almost impossible (and at the same time would require an impressive effort). Hence, they pursue a methodology for integrating and reusing existing human resources management standards like the Statistical Classification of Economic Activities in the European Community (NACE)[21], the International Standard Classification of Occupations (ISCO)[22] and the International Standard Classification of Education (ISCED)[23], among others. The results achieved with the above methodology are again integrated in the HR-XML ontology, in the form of a number of sub-ontologies dealing with information about job seekers, job offers, compensation, driving license, economic activities, occupations, educations, geography, labour regulations, languages, time, skills, competency, etc. This way, a reference ontology is actually created, playing the role of the common language enabling systems seamless integration. At the same time, ontologies exploited in the various employment services can be mapped to the reference ontology and each local system is allowed to use a vocabulary suited to the particular market.

It is worth observing that in some cases, the objective of finding the best match between a job demand and a job offer is shared with the symmetrical or complementary issue of finding the right expertise, i.e., the right expert, for filling in a given competency lack. Therefore, a number of works dealt with the problem of recommending the right candidate for a given position. In this case, the difficulty

[21] http://epp.eurostat.ec.europa.eu/cache/ITY_OFFPUB/KS-RA-07-015/EN/KS-RA-07-015-EN.PDF

[22] http://www.ilo.org/public/english/bureau/stat/isco/index.htm

[23] http://www.uis.unesco.org/Education/Pages/international-standard-classification-of-education.aspx

is especially in properly defining the concept of *expertise*. Expertise is highly dynamic, difficult to measure and variable in strength [Fazel-Zarandi et al. 2011a]. Like in previous works analysed, experts profiles are often considered as collection of competences or skills (either declared by the individual or inferred from various kind of sources, including social media [Hansen et al. 2010]) and the key issue is again about how to describe data involved with a shared vocabulary, how to deal with missing or incomplete information, how to exploit available measures to come out with a final similarity score to be used for performing the required comparisons, rankings, etc. A similar goal is pursued in [Reich et al. 2002], where the authors address the formalization issues related to the development of a semantic solution for finding people with a certain skills profile to be used for staffing new projects, identifying experts who might help to solve a certain problem, etc. Some of the topics discussed above have been considered also from other perspectives, e.g., with the aim of formalizing education systems, course syllabi, etc. [Koper, 2004] [Cubillos et al. 2008] [Cubillos et al. 2007b].

Independent of the actual application for the outcomes of the knowledge representation step described above, the review of works considered clearly showed the intensity of the human effort associated with the creation and maintenance of the ontological models required. In some cases, the overall complexity might be somehow relaxed by exploiting some statistical, clustering, pattern recognition or fuzzy logic methods that could be used to let the machines learn what the relevant concepts and relations between concepts are, rather than relying on articulated ontological models managed by human experts. Nonetheless, as it will be shown in the following, a strong trade-off is usually observed between the accuracy of the final application results and the quality and completeness of the underlying formalization and annotation steps.

Annotating, i.e., extracting knowledge from, résumés and job offers

Indeed, the final success of an online job matchmaking system is related to the effectiveness of the matching and ranking algorithms adopted. However, like with any automatic information processing tool, these algorithms definitely need a solid knowledge base to work with, represented – in the case of semantic processing – by annotated documents generated by job seekers and hiring departments. As discussed above, résumés and job offers can be characterized by a variable shape, which could range from fully structured to completely unstructured. Despite such heterogeneity, data contained therein can be roughly classified in several sections dealing with education paths followed and titles achieved, past employers and positions covered, capabilities and personal attitudes possessed, etc. In particular, prior experience is generally described by making reference to a set of competences or skills that, according to the literature review, have been modelled in many different ways but, in any case, sticking to taxonomies, ontologies, i.e., in one word, to a shared vocabulary.

Unfortunately, the effort of annotating a document based on a some formal model in a fully manual way could be at least comparable to that of constructing the model itself [Fazel-Zarandi et al. 2013]. Time and cost benefits expected from the application of semantic techniques would be therefore definitely ruined without a practicable way for easing such knowledge elicitation step [Brown, 2004]. Hence, totally or partially automated techniques for matchmaking data extraction have received particular attention of both researchers and human resources management companies in the last years [Mochol et al. 2007a] [Colucci et al. 2003].

On the one side, an automatic résumé or job offer parser should be able to break collected documents in their composing parts and extract relevant information from them, by linking words found to the underlying concepts and context, thus creating a structured picture of the person or the open position concerned. For instance, some of the available job portals advertise the fact that their semantic engines are able to post-process collected data to identify titles and employers, to automatically determine the years of experience with particular skills, etc.

On the other side, recent experiences have demonstrated that the best results are obtained by mixing post-processing automatic activities with a pre-processing stage where end-users actively participate in the annotation step [Bourse et al. 2002]. In this respect, it is worth saying that the employment market is gradually adopting ever more semi-structured forms for acquiring user data [Bizer et al. 2005] [Ionescu et al. 2012] [Rafter and Smyth, 2001] [Kessler et al. 2008]. The goal is therefore to keep the user interface very easy and far from the underlying technology, in order not to request job seekers and employers to reason in terms of ontologies, semantic relations, sophisticated query languages [Brown, 2004] [Bernstein et al. 2004], etc. More in general, as underlined in [Colucci et al. 2006], the goal should be to directly involve the end-users in the overall search process and to support them in the seamless achievement of the final results expected. In particular, a semantic system should ensure efficiency and trust. That is, it would important to find the right matches best fitting users needs. Furthermore, users should be confident that the system has made the best choice for them. A simple presentation of results may be not sufficient in this sense and suitable explanations/motivations should therefore be provided to convince them about the effectiveness of the particular approach pursued. Also, the system should let the users experiment different criteria for sorting our results.

Taking into account the above observations, platforms developed could be also distinguished based on the user friendliness of the interface made available for letting job seekers and hiring departments publish their curriculum vitae or open job positions. Unfortunately, when research works are considered, it can be immediately realized that the amount of attention paid to application usability is generally inversely proportional to the algorithmic depth of the approach presented. That is, when the focus is on representing knowledge or solving the matchmaking algorithmic problem, the degree of friendliness of user interaction is often disregarded. To make an example, in [Yahiaoui et al. 2006] a specific domain ontology is created and a competency-based semantic matching mechanism is defined. However, annotation is expected to be carried out by the end-user by directly working with an ontology editor like Protégé operating directly with the low-level languages of the Semantic Web. In other cases, the presence of a framework supporting end-users in the annotation step (or, more in general, of a user interface) is only theorized [Lv and Zhu, 2006] [Bizer et al. 2005] [Fazel-Zarandi and Fox, 2009] [De Coi et al. 2007] [Fazel-Zarandi and Fox, 2010] [Ionescu et al. 2012], without making reference to a concrete implementation or, in some cases, saying that it is still under development.

When a easier interface is available, it generally requests the individuals to provide accurate and comprehensive descriptions of their competency [Earl, 2001], possibly with the support of some facilitation system. For instance, in [Vega, 1990] a simplified approach based on the use of natural language is exploited. In this case, a concrete tool with an ad hoc user interface was developed, though dedicated solely to job seekers. Candidates interact with a text-only template designed to let them insert the curriculum vitae. The system identifies unknown words (e.g., due to typing errors and spelling mistakes), allowing for the creation of an annotation which is aligned with the underlying linguistic, employment as well as job offer-related databases and knowledge bases. In [Reich et al. 2002], individuals specific their skills by selecting concepts from a terminology and by indicating a level for each selected skill (from elementary to expert). The system also manages information about education and job functions.

In the framework designed by the authors of [Bourse et al. 2002] [Radevski and Trichet, 2006] [Trichet, 2004], a number of competency ontologies (related to a sector, enterprise or aimed at modelling personal attitudes, etc.) are expected to be exploited in a annotation framework directly involving the end-user. In the authors' vision, when writing a curriculum vitae it is usually difficult to choose the best sentences (in natural language) for expressing the competencies acquired during a professional history. Sometimes, the sentences adopted are not particularly significant and no not include or reflect in a precise way the individual's profile. Thus, having a reference of the tasks and competences underling a profession can be an important help when dealing with identification of the a candidate's competences. The same reference can also be used also by the job seeker to evaluate whether owned competences are compatible with those requested by a given profession. The idea experimented is therefore to first let the user choose a sector from a pre-defined hierarchy. Then, the profession of the sector that better characterizes his or her work experience is chosen. This way, a number of tasks or missions are displayed, linked to the particular profession. In the ontology, each activities is associated with competencies mobilized (and related resources, i.e., knowledge, know-how and behaviour). He or she could then add details about the context in which the particular competency has been performed (considering social and organizational aspects, economical constraints, physical characteristics, informational details, etc.).

A similar approach based on repeated choices is illustrated in [Colucci et al. 2006], where the authors present a totally visual approach based on query refinement. In the specific case, the approach is exploited in a slightly different application scenario, though the focus is still on recommendation and matchmaking. The visual interface developed moves from the idea of supporting the user while he or she is working on general queries that are progressively refined by exploiting the ontological model beneath.

The issue of directly producing an annotated résumé is explored in [Mirizzi et al. 2009]. In this work, domain knowledge is used both to support the individual during the elicitation of his or her abilities during the composition of the curriculum vitae as well as for the final presentation of the annotated curriculum vitae itself. A semantically-guided interface is developed, that is able to suggest concepts to be used in the annotation by means of a tag cloud [Mirizzi et al. 2010] made up by exploiting the relations between competency elements. Basically, the approach is based on an initial semantic-driven tagging step. In the authors' vision, this should allow both to overcome the limitations of existing web-based systems for résumé writing as well as to make more accurate the next search steps. That is, before starting to write down in natural language text about his or her working experience, the job seeker is asked to collect his or her skills and competences as a set of tags. Based on typed keywords, the system recommends [Adomavicius and Tuzhilin, 2005] a set of possible tags which are semantically related to such keywords (based on keywords completion and on an ontology that, in the specific case, is related to the ICT domain, as well as by considering also the part of the résumé that has already been created). Size of tags indicates the relevancy of terms. The job seeker can select the tags that better suits his or her needs. Tags are added to a so called tag bag. Once a tag bag has been populated with all the needed tags, he or she can start filling in the forms devised for compiling the curriculum vitae, thus describing relevant work or education and training experiences. Then, he or she could pick up tags in the bag to annotate such experiences.

According to the review of works above, a common trend consists in directly involving the user in the annotation of documents handled by semantic-based tools, including those developed in the frame of e-job seeking and e-recruitment processes. However, it has to be underlined that, both when knowledge extraction is performed as post-processing (for automatic parsing) or a pre-processing (for facilitating the interaction) step, the final effectiveness of a particular technique is given, on the one side, by the quality of the annotation produced and, on the other side, on the accuracy of the semantic algorithms adopted, which should be able to mimic human reasoning and precisely measure the distance between words and sentences based on their meaning. Thus, in the following, the approaches developed to deal with such issues will be considered in details.

Matching and ranking approaches: comparing descriptions based on semantic similarity

The final step of the online matchmaking process being discussed is well represented by the actual matching phase, where achievements and attitudes described in job seekers' résumés are compared with job offer requirements and constraints. At a very high level of details, the result of such comparison is, depending on the perspective considered, a list of job postings or of candidates. The list should report in the top positions the best matches achieved. That is, it is not just a matter of matching, but also of ranking. Being almost all based on ontological annotations defined starting from the common core concepts of taxonomy and ontology (though defined in a possibly heterogeneous way), the various approaches proposed differ in the algorithms used for exploring the hierarchical collection of concepts, for navigating relations among concept nodes, and for finally taking into account possibly weights and/or constraints defined by the users. In one word, in the approach pursued for defining and computing *semantic similarity* [Lv and Zhu, 2010].

In [Vega, 1990], a platform which is in operation since 1989 and that has been developed to support job seekers in the process of consulting job postings is presented. Though at that time the Semantic Web vision had not been formally expressed yet, the authors of [Vega, 1990] illustrates the design of a so called matching machine, a semantic algorithm capable of extracting normalized terms from the user's request of curriculum vitae, of associating extracted information to the corresponding concepts in the knowledge base, of spreading the initial concepts over the knowledge base relations and of finally extracting and classifying indexed job offers based on the initial and spread knowledge. In [Yahiaoui et al. 2006], the matching algorithm considers weighted hierarchies of knowledge elements related to the ICT field. When the requested knowledge element matches the demand, the weight is cumulated. When an exact match is not found, the tree structure is navigated to determine an adjusted weighted contribution for similar information. That is, as anticipated, semantics is exploited to let the matching algorithm manage the variable completeness/incompleteness of descriptions adopted by end-users.

In [Lv and Zhu, 2006], besides focusing on the definition of an ontology represented by a skills graph, the authors propose to measure semantic similarity between résumés and employer's requirements by determining the shortest path between concepts in the ontology and by taking into account edge weights. In the above process, weights are also considered with respect to qualitative thresholds set by the employer on multiple skills possibly needed, which are expressed in the form of a priority score. In [Lv and Zhu, 2010], the authors of [Lv and Zhu, 2006] extend the proposed approach by presenting a matching method that is based on demand constraints. The method tries to mimic the behaviour of the job seeker or the company, which do not make restrictions arbitrarily to the respective offer. That is, in the devised technique demand constraints can be either rigid (about gender, education level, graduate school) or elastic (pertaining work experience, professional knowledge, skills, job potential, etc.).

In [Fazel-Zarandi and Fox, 2009] a methodology for taking into account descriptions mentioning both required and preferred requirements is presented. The strategy considers competences (expressed as skills and proficiency levels) associated to both work experiences as well as formal and informal learning. A similarity-based ranking is then proposed, working with both under-qualified applicants as well as considering nice-to-have requirements. Something very similar to the situation of [Lv and Zhu, 2006] and [Fazel-Zarandi and Fox, 2009] occurs also in [Bizer et al. 2005]. Here, the authors design a way for integrating a wide range of existing taxonomies, ontologies and classifications with newly defined models into a wider human resources ontology serving as shared formalization for dealing with skills, occupations, industry sectors, etc. Then, they coupled the above results with a semantic matching algorithm able to combine annotations using controlled vocabularies with background knowledge of the specific domain to determine the similarity among concepts, which basically constitute the building block for the creation of a ranked list of alternatives (either job offers of job applicants). They use again a measure based on the distance between concepts in the hierarchical models [Billig and Sandkuhl, 2002] [Zhong et al. 2002]. In addition, they explore the concepts hierarchy under the assumption that two general concepts are less similar than two specialized ones by also keeping into account the competence level (expert, novice, etc.) and the importance of specific requirements. An overall similarity formula is designed, which is based on weights computed as illustrated above.

In [Mochol et al. 2007a], the authors of [Bizer et al. 2005] combine the approach considered above with the concept of query relaxation to let the system deliver responses even when a complete match of the acquirements/requirements cannot be achieved, though results could be nonetheless acceptable for the employer or the job seeker, respectively. The approach consists in first checking data by using the strongest possible constraints (which should return the best answers satisfying most of the given conditions) and then weaken the query if the returned set is empty or does not contain satisfactory information. In [Mochol et al. 2007b] and [Mochol et al. 2007c], a similar concept of query approximation is used. The idea is again that of letting the user operate on the interface for performing the query by changing some constraints. A query rewriter then transforms the new requirements in a smoothed request that is exploited for expanding the search space. In [Fazel-Zarandi and Fox, 2009] and [Colucci et al. 2003], Different kinds of matchmaking strategies over description logic-based models are combined to specifically deal with under-qualified or incomplete matches and rankings.

After having considered the key steps involved in the semantic processing of job seeker and employer-related information through a comprehensive review of existing literature, an overview of commercial platforms available is reported. Non-semantic solutions are considered together with semantic ones, though – as it will be shown – even for established solutions the trend is towards a ever more significant exploitation or integration of semantic technologies.

2.4.2 The business side of job matchmaking: commercial web platforms available

As seen while reviewing the academic literature, technology is strongly influencing the way hiring activities are carried out. In fact, if in the past, sending a job application by mail or e-mail was a praxis, and the manual analysis of curriculum vitae was a time-consuming activity, today more and more commercial instruments reducing barriers between companies and people looking for a job and supporting the evaluation of candidates' résumés have been developed. As anticipated, semantic approaches are progressively replacing solutions based on alternative approaches. However, it is interesting to review all the alternatives before coming to a final comparison.

In this perspective, a first meeting point between job seekers and employers is represented by job boards, websites allowing companies or recruitment agencies to publish (many times paid) job postings in order to advertise open positions. These platforms, such as CareerBuilder[24] and Monster[25], guarantee to companies willing to pay for a job advertisement, a significant visibility, due to the huge number of users browsing their pages. Job seekers can first perform job offers searches by specifying the job title or the required skills, and, then, can apply for a job.

In parallel to job boards, few years ago, job search engines and job offers aggregators like Jobrapido[26] or Indeed[27], started to appear on the web, in order to answer the need of simplicity and speed, when looking for a job. The idea underlying a job search engine is that the user (the job seeker) has no more to trawl through different job boards every day in order to monitor new job offers, since the aggregator is able to crawl and index job offers published on different websites (e.g., job boards, corporate web pages, etc.), to filter them on the basis of the professional figure or

[24]www.careerbuilder.com

[25]www.monster.com

[26]www.jobrapido.com

[27]www.indeed.com

the geographical location, and to eventually notify the job seeker each time a new job offer that could potentially interest him or her has been posted.

The collection and the indexing of job offers coming from the whole web could be useful not only for providing job seekers with a unique platform for performing job searches, but also for carrying out statistical analysis on supply and demand. This is the idea behind Wanted Analytics[28], a real-time business intelligence tool for the talents marketplace, which collects, besides online job postings, additional data from information sources like financial markets, governmental statistics, etc. After the computation of the gap between the supply of workers and the local demand for workers, the website provides job seekers and employers with instruments allowing them to take more accurate choices, such as, for example, the analysis of how easy is getting employed/hiring staff or what is the average wage in a given region or professional field.

If, on the one hand, the search for a job/employee has been considered until few years ago as an activity to be carried out individually, for which the company website could be at most the interface between job seekers and employers, today social networks play a more and more important role in the selection of the job offers to apply for/the candidates to interview. A couple of examples in this direction are Glassdoor[29] and InsideTrak[30], job search engines merging job offers with information from a community of users. Besides receiving a list of job offers based on the professional figure and the location specified, a job seeker exploiting these websites can read the reviews of other users that previously worked in the various companies reported, receive information about the average salary, or even see the pictures of the offices and get hints about the more frequent questions asked in the interviews.

The importance of networking had been already foreseen in the first years of 2000 by Facebook[31] and LinkedIn[32]: today, these social networks count 1 billion and 175 million users respectively. However, while, for years, LinkedIn has been the leader of work-related social networks, a number of applications supporting human resources recruitment was recently developed by Facebook: one example is Social Jobs Partnership (SJP)[33], a tool allowing recruiters to share open positions with the Facebook community. From the SJP search page, each Facebook user can browse job offers coming from job boards such as Monster or US.jobs, send a message

[28]www.wantedanalytics.com

[29]www.glassdoor.com

[30]www.insidetrak.com.au

[31]www.facebook.com

[32]www.linkedin.com

[33]www.facebook.com/socialjobs

containing the job offer to his or her friends, or “like” it. MyParichay[34], another Facebook application, is rapidly spreading over India: here, companies can post their job offers for free, and reach more than 4.5 million professionals. Job seekers, on the other hand, can create a professional profile, search for jobs coming from job boards as well as from Facebook pages, and directly apply, or see, if among their contacts, there is someone working for the company that is publishing the job advertisement. And it is the focus on social networks that led up to the development of applications such as BranchOut[35] and BeKnown[36]: these tools allow the users to create, in parallel to the private Facebook profile, a work profile, to share with their network interesting job offers, to identify people working in the company of their dreams in order to add them to their contacts. In a similar way, the Jackalope Jobs[37] platform intensifies the exploitation of social connections, by combining job search and social networking: here, the job seeker can see, together with a number of job offers coming from job boards as well as from corporate sites answering his or her query, which people, among his or her contacts, could open the doors for working in a company, simply because they already work(ed) there and, thus, could endorse him or her.

However, if knowing someone inside a firm could increase the probability to be hired, it is worth remarking that a considerable part of the hiring process is based upon the analysis of candidate’s competences and skills, by matching them with the job offer. To this extent, LinkedIn provided its users with new features, such as Skills and Expertise[38]. This instrument allows the job seeker wanting to work in a given context to identify which are the requirements. By searching for a skill or an expertise, he or she is provided with a list of skills related to it. This way, he or she can find what is lacking in his or her professional life, in order to enhance his or her career on the basis of the requirements of the world of work, or of curriculum vitae wrote by other LinkedIn users. Moreover, it is also possible to see skills trends, together with more relevant companies that hire people with the above characteristics, or professionals having a given skill and the type of relation they have with the user. On the one side, this feature provides job seekers with an instrument for planning their personal and career growth and for making their competences and skills explicit. On the other side, it gives companies a tool for identifying industry terms they might overlook while writing a job advertisement.

Other websites already made the computation of the match between job offers

[34] apps.facebook.com/myparichay

[35] branchout.com

[36] www.beknown.com

[37] www.jackalopejobs.com

[38] www.linkedin.com/skills

and demands their key strength, specifically combined with the use of semantic-based approaches. An example is Bright, which developed the Bright Score Calculator[39], a tool helping recruiters to match résumés with job descriptions. Companies can cut and paste their job description, upload candidates' curriculum vitae and receive results on the basis on how suitable a job applicant is for the job offer. For the computation of the match, together with keywords this tool considers also synonyms, employment history, education, etc. Such a tool, exploiting machine learning algorithms based on the analysis of 2.1 million job descriptions and 2.8 million résumés, could also help job seekers in finding their rank with respect to a job offer. In fact, job seekers could insert a job advertisement, upload their curriculum vitae and get as a result a score representing how much they are attractive for a given company. Yet, search engines such as Ask[40], Bing[41], Google[42], Yahoo[43] etc. already exploited combinations of lexical, ontological, statistical, user-behavioral learning and, recently, even semantic technologies in an attempt to better understand what searchers, possibly job seeker and employers, really want, instead of simply providing an exact match for the keywords typed. Also Monster carried out a relevant investment for the creation of tools allowing to effectively search for potential candidates for a given job. In particular, in its Power Resume Search[44], Monster simplified the analysis of a huge number of résumés in a short time by using semantic technologies. With this tool, firms can search for a job applicant by simply inserting the required job title, since the system is able to find professionals which are similar to the requested one and to exclude from the list those entries not satisfying the requirements that would be nonetheless returned by a pure keyword-based search. As a consequence of the introduction of semantic technologies, companies no more have to conduct complicated boolean searches, with the aim of including all the possible combination of terms, or for leaving out concepts that could erroneously influence the match. Moreover, employers can further refine the search, e.g., on the basis of the years of experience or of the skills required or nice-to-have for a given job. The result is an ordered list of scored candidates, with an highlight of matching characteristics. Another solution, designed for recruiters and applying semantic technology to search for résumés is TalentSpring[45]. Here, structured information

[39] www.bright.com/score-calculator

[40] www.ask.com

[41] www.bing.com

[42] www.google.com

[43] www.yahoo.com

[44] hiring.monster.com/jcm/resumesearch/resumesearchtestdrive.aspx

[45] www.crunchbase.com/company/talentspring

such as people profiles on social networks is also considered and compared to employment requirements using both ontological categorization and semantic analysis. Like the above platforms, Jobnetchannel[46], a channel for job recruitment, provides companies with an instrument for matching job offers and demands not only on the basis of a keyword-based comparison, but also by exploiting semantics to perform reasoning on inserted résumés. Here, the job seeker has to follow a guided path for the creation of his curriculum vitae, which allows him or her, among other things, to make skills achieved in formal, non-formal and informal contexts explicit. The computation of the match is based upon four parameters: number of training years, evaluation, use and regularity of each skill, related to other skills by means of a taxonomy. While the above websites mainly deal with matching candidates' competences and skills, together with the working experience, Path.to[47] aims at finding the best combination job seeker – company by taking into account also other aspects of the working dimensions, like location, salary, benefits, work environment and dress code, as well as ideal company size. Starting from the user's profile, this platform drafts a list of ranked job offers to apply for. In addition, a weighted endorsement mechanism based on social networks and professional communities is considered, in order to compute a better match. Moreover, in this view, both employers and job seekers can like or dislike the results, thus improving future results.

2.4.3 Final consideration and remarks

Indeed, the above discussion, that reviewed the bulk of the literature about automatic (and, specifically, semantic) job matchmaking, should be enough for convincing the reader about benefits that could be expected from the implementation of computer-based frameworks for supporting job seeking and job recruiting processes. Similarly, the role and importance of semantic techniques should have been clarified and demonstrated as well by means of a significant number of study cases.

However, it is worth observing that opposite to the above advantages, several risks and difficulties in the usage of semantic solutions for job matchmaking have been also identified. The main peril is of trusting the power of semantics too much, and work with matches obtained with terms that were certainly related, though not necessarily relevant to the search, to the domain, etc. Ironically, a résumé crammed with annotations could be preferred to a synthetic and organized one because of the amount of semantic matches. Moreover, semantic engines generally interface with the end-user by providing a ranked list of results. Indeed in some cases ranking scores are accompanied by a kind of explanation. Nonetheless, in the

[46] www.jobnetchannel.com

[47] www.path.to

solutions developed until now, the experience of a recruiter or the judgement of the job seeker can re-enter the process only at this pre-negotiation phase. However, in the previous processing step, interesting possibilities could have been discarded and, hence, lost. Then, there exists limitations that are not related to the use of semantics in the job matchmaking domain only, but to online recruitment in general. For instance, it has been demonstrated that in the last years, some of companies that initially invested significantly in online recruitment have now shifted back to traditional recruitment processes [Lee, 2011]. This is at least partially due to the fact that, even though the web has opened new powerful opportunities for job seekers (who can now quickly post their résumés and apply for many job positions) and companies (that have got much more applicants for a given job offer), the amount of under-qualified applications has grows tremendously. The consequence has been that a technique eventually capable of speeding up and enhancing the quality of the selection process has been used to process a greater amount of data in a comparable time with the risk of missing talents and right candidates in the mass. Finally, in a ideal world where the perfect platform has been created, all the companies and job would probably use that platform: unfortunately, all the job seekers and employers would be presented the same best matches with the consequence of an explosion of the following negotiation process. Paradoxically, talents not publishing their profile online or companies opting for traditional recruitment would still remain unmatched.

Hence, the real contribution of automatic and semantic-based approaches not only to the job matchmaking domain but rather to the overall policies set for broad domains like the education and training and the labour ones as well as to priorities established in terms of market transparency, mobility, inclusion and the like, could only be achieved and sustained if all the core elements in the overall semantic job matchmaking process will be properly setup. Indeed, some of the pillars required for the implementation of the global picture described are in place already since some years [Vega, 1990]. For instance, as shown, semantic job portals where résumés and open positions can be posted to and searched for are no more a novelty. However, the final success of such solutions will depend on how the core challenges associated with the evolution of enabling technologies in the specific domain of semantic job matchmaking will be addressed.

As said, one key challenge will be about how to create, maintain and evolve truly comprehensive ontological descriptions (especially the competency one) able to go beyond the boundaries of a particular academic experience or company product. To make an example, since the search space of semantic engines is defined by the breadth of the underlying knowledge model, a particular acquirement or requirement could remain unmatched because a relation between two terms referring to the same concept or to similar concepts has not been foreseen during the ontology definition or updating steps. In fact, on the one hand, developing a domain ontology and achieving a wide consensus in a broad domain could be tasks hard to accomplish.

On the other hand, once developed, an ontology preserve its usefulness only if it is able to keep up with the dynamic nature of the surrounding context (changing needs from the labour word, changing learning outcomes in education path, etc.). Therefore, though the interesting efforts being put on the automatic generation of domain models, a special attention will have to be devoted to the identification of suitable strategies for supporting the collaboration of knowledge engineers on the management of ever more complete and integrated formal representations of the contexts of interest, which should be able to incorporate the outcomes of relevant previous experiences and standardization initiatives [Fazel-Zarandi et al. 2013] [Brown, 2004].

Then, another key challenge will be represented by annotation. As illustrated, what it is often critical in existing implementations is to link user-defined data with the concepts defined in the shared semantic repositories, i.e., to extract and relate to terms and sentences the meaning that was originally expressed by the job seekers or the employers while drafting their résumés or job postings. To make an example, an annotated job offer with only a few requirements matching with the acquirements mentioned in a given résumé could be preferred by a semantic engine to a job offer where all the requirements are met but annotation is lacking. Hence, would it be possible for a job seeker or employer to describe, depending on the perspective, acquirements or requirements in a general way without being requested to pick up definitions from a pre-defined list of occupations, knowledge items, skills, etc. (and possibly without sticking to a specific template and having to publish data in a number of dedicated not-communicating services), then the usage of such solutions would become really seamless and a ever more widespread diffusion of semantic-based tools could be expected. To achieve such goals, annotation needs to be automatic and it must not require the users to go around tagging and cataloguing content in order to make it acceptable for computers to understand. A special focus would have to be devoted in particular at simplifying and improving the power of user interfaces devised [Brown, 2004]. By the way, though annotation could be regarded as a time consuming operation, in semantic processing being able to get relevant results could be also a matter of time [Lv and Zhu, 2006]. In fact, when it comes to disambiguate among the various meaning a given term can have, the number of relations can be so high that the number of possibilities to be explored could be so large that the necessary computing power, or time, could even be not sufficient (or not adequate for the problem addressed). Moreover, linked to time, there is the issue of quality of results produced. In both cases, it is a matter of properly exploiting the formalization and annotation results, which is actually the last main challenge to be considered. In fact, a huge number of algorithms have been experimented already, even in the context of public platforms. In most of the cases they were tailored to the specific knowledge base underneath. Nonetheless, most of the approaches adopted could be possibly generalized to deal with other knowledge models available or still to come. Moreover, a number of generic approaches for measuring (semantic) distances among concepts are continuously proposed in the literature [Fazel-Zarandi and Fox, 2011b]. In general, they are not exquisitely designed having in mind the specificness of the job matchmaking domain, but also in this case they could be easily adapted to it. However, it is worth observing that the way semantic approaches, like for instance search or recommendation systems, are generally evaluated is by means of precision and recall measures. Unfortunately, in the particular domain of interest, experimental studies are in most of the case limited to the area of academic research. A rare example is represented by [Lv and Zhu, 2010], where the authors carry out an evaluation on their proposed matching approach by simulating a recruitment process where a firm needs to recruit a job seeker with a specific profile. After having pre-collected a number of résumés, human resources managers from a professional company in the sector concerned are invited by the authors to carry out the manual selection process. Then, they compare the top candidates pre-selected by the platform with those resulting from the human evaluation and they observe a 90% accuracy in the results. Indeed, more data are available for public platforms, but in this case it is more difficult to measure the effectiveness or get technical details of the approaches adopted. Hence, a diffusion of these techniques will additionally contribute at providing more study cases to be then assessed from the point of view of all the stakeholders involved. Nonetheless, as reported in the future activities section of many works in the literature, significant efforts will have to be devoted in particular to check the performance of the various approaches proposed and select those that could be elected as the best choices for a given application scenario or that could be considered as a basis for the construction of further evolved solutions. Based on such results, ad hoc business models could then be designed, possibly evolving from the present ones [Lee, 2011], where, for instance, e-job seeking and e-recruiting could be strongly integrated with the overall human resources management frameworks [Bizer et al. 2005] [Bollegala et al. 2011].

Chapter 3

Using taxonomies to support the construction and comparison of qualifications

3.1 Introduction

During last years, issues related to mobility of students and workers across Europe gained more and more relevance: nowadays people have better chances to spend a studying or working period abroad to acquire missing competences, or to find better working opportunities. However, marked differences in the meaning, content and interpretation of tasks and functions as well as of learning outcomes in the framework of the European labor market and educational offer still limit the mobility of workers and learners.

In this context, the correct identification of competences fulfilling training gaps, as well as the depiction of acquired competences (according to a shared formalism) becomes of primary importance for an individual.

For this reason, since the establishment of the European Union, several instruments have been developed in order to guarantee comparability, transferability and recognition of qualifications across different countries, as well as to enhance transparency and mutual understanding across Member States.

One of the most relevant tools is represented by the EQF. According to the EQF, lifelong learning qualifications are categorized in eight reference levels (from one, the lowest, to eight, the highest), and associated learning outcomes are described in terms of knowledge, skills and competence concepts, thus opening the way for the creation of a shared understanding in the lifelong learning domain.

Thanks to the EQF, qualifications (and curricula, as well) described according to the above guidelines have more possibilities of being understood and referenced by all the actors involved in the educational and occupational domains. Nevertheless, when it is needed to compare education and training contents as well as personal abilities for mobility purposes, a mere application of EQF principles (e.g. carried out on manual basis) may not be feasible, since a huge amount of information has to be considered. Hence, ad hoc instruments, able to semantically compare information embedded into qualifications, course syllabi, personal resumes, etc. by exploiting the EQF guidelines, should be created.

The objective of this Chapter is to present the methodology developed during the Ph.D., within the TIPTOE "*Testing and Implementing EQF and ECVET Principles in Trade Organizations and Education*" project. Politecnico di Torino was partner of this project, and part of the Ph.D. research activities have been carried out to address the above need. In a nutshell, the TIPTOE project aimed at lowering barriers between the labor market and the training dimension at the European level, by identifying a common European profile in a specific scenario (in the particular case represented by the trade sector) to be exploited for experimenting EQF-based automatic processing of relevant information in the mobility perspective. As it will be detailed in the following, the development of a semantic-based engine able to perform an EQF-aware taxonomy-based comparison of country-based formative offers (expressed through educational profiles) and labor market requirements (represented by national occupational profiles) with the goal of finding similarities and specificities emerging from heterogeneous "local" descriptions structured in terms of learning outcomes played a key role for both identifying the common profile and for supporting the deployment of services targeted to end-users.

The rest of the Chapter is organized as follows: Section 3.2 explains the objectives of the TIPTOE project. In Section 3.3, the four-step methodology developed for the creation of the common profile - *information collection*, *taxonomy construction*, *definition of inference rules and approaches for semantic comparison* and *common profile creation* - is presented. Finally, Section 3.4 illustrates several services (the *automatic identification of the EQF level* and the *EQF ruler*) targeted to end users, whereas conclusions are drawn in Section 3.5.

3.2 The TIPTOE project

Mobility of individuals can be an instrument to address existing skill shortages and mismatches in a country or region, thus improving the efficiency of the labor market and removing brakes on economic growth. However, even though all the parties involved could benefit from transnational mobility, there are still several barriers to this process, mostly due to the differences in the meaning, content and interpretation of tasks, functions and learning outcomes to be carried out by and expected by European workers and students.

For this reason, the objective of the TIPTOE project is to tackle the problem of interpretation and application of EQF principles from a practical perspective, by specifically focusing on a sector that is considered of primary importance (that is the trade one) and by proposing a methodology capable of mitigating both the gap between the different European education and training systems as well as the (cultural) differences in the content and interpretation of occupations within the European labor market.

The basic assumption behind the TIPTOE project is that labor market and educational field have both their own understanding of which kind of knowledge, skills and competence are related to a given professional; nonetheless, frequently, qualifications provided by education and training institutions do not really reflect labor market's needs. Moreover, the lack for rules outlining a minimum set of knowledge, skills and competences that a student should possess at the end of a training path creates strong information asymmetries between the education and the labor worlds and severely limits the mobility among countries.

In fact, since a unique and well-defined qualification profile is missing, employers may ignore the exact contents of the courses attended by a student who is applying for a given job position, and consequently, may not know which learning outcomes he or she actually achieved. The depicted scenario is even more jeopardized and complex in a transnational perspective, especially when non-formal and informal learning paths are also taken into account.

Within the TIPTOE project, both the labor market and the training field are investigated and compared at European level, in order to identify the core elements (knowledge, skills and competence) characterizing a professional, which are then exploited to build a European-wide profile. The strategy pursued throughout the project strongly relies on semantic instruments for allowing an automatic and syntax-independent comparison.

One of the main outcomes of the project is the development of a web portal supporting users in the creation of the common European profile, and providing stakeholders of the trade sector with a set of services (exploiting the knowledge base developed throughout the project) aimed at the identification of the EQF level of a qualification owned by a worker or provided by a training institute.

For what it concerns the creation of the common European profile, the methodology adopted throughout the project is the following: first, a series of interviews with relevant stakeholders of the labor world have been carried out. The objective of this first phase is to outline a set of tasks a worker should be able to accomplish by characterizing them in terms of knowledge, skills and competences. Secondly, an investigation on the education and training field is performed: in this phase, several interviews with relevant training organisms are conducted in order to identify which learning outcomes a student should possess at the end of a formal training path. After this phase, the two sets of information are to be compared in order to identify the common elements between the requirements of the occupational domain and the outputs of the educational routes: this comparison aims at defining a unique profile, drawn according to the EQF principles. Then, in order to catalogue the outcomes of the interviews in a structured way and to perform the required semantic reasoning onto them (thus avoiding possibly incorrect results provided by a manual comparison carried out on a huge amount of data), elements belonging to occupational and educational descriptions have been linked to a set of concepts, organized into a taxonomy: as a result, relations among elements of the descriptions and concepts of the taxonomy could then be used to carry out the necessary reasoning and, thus, overcoming lexical barriers.

Based on the outcomes of the above phases, several services specifically addressed to project stakeholders have been developed. In particular, in order to support users in the identification of the EQF level for a given qualification, a specific tool for the automatic referencing has been implemented. Moreover, results provided by the semantic engine have been collected and structured according to their EQF level in a *EQF ruler*, so that to provide a quick and easy-to-read representation of the common profile.

3.3 The methodology

The methodology devised creation of the common European profile consists of four stages: *information collection*, *taxonomy construction*, *definition of inference rules and approaches for semantic comparison* and, finally, *common profile creation*. In the following, each stage will be discussed in details by making reference to Figure 3.1, where the overall methodology is summarized.

3.3.1 Information collection

As already mentioned, the *information collection* stage was carried out in order to collect the requirements of the labor world and the outputs of the education and training domain, expressed in terms of units, task, subtasks as well as of knowledge, skill and competence elements.

In order to define a shared format for collecting information (and then representing it in the taxonomy construction stage) the representation of knowledge, skill and competence concepts made by [Pernici et al. 2006], and already presented in Chapter 2 was exploited. As already said, according to [Pernici et al. 2006], a knowledge could be defined as a set of knowledge objects (KO), a skill could be represented as a KO “put into action” through an action verb (AV), hence by one or more pairs KO – AV, and a competence could be identified as a triple KO – AV – CX, that describes the ability of putting into action a given KO in a specific context (CX).

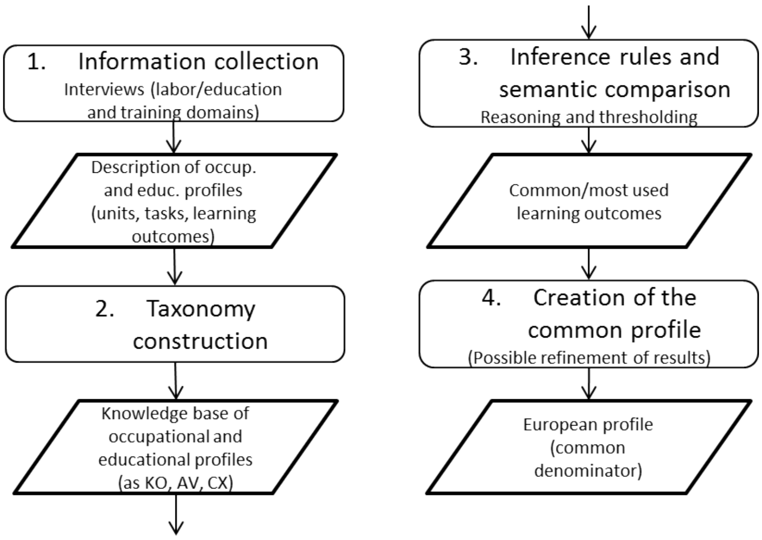

Figure 3.1. Overall methodology for the creation of the common profile

The relations among the above concepts have been taken into account during the information collection phase.

In particular, four relevant professional profiles were identified in the selected sector (namely Shop Assistant, Shop Manager, Logistic Assistant and Logistic Manager), and interviews were conducted on both the working and the training contexts in order to identify their key elements. At a first stage, stakeholders (i.e. employers of the retail and wholesale sectors) belonging to the labor dimension of different European countries (i.e. France, United Kingdom, the Netherlands, Italy, Lithuania, Portugal, Germany and Slovenia) were interviewed in order to collect, for each of the four profiles, a list of knowledge, skills and competences that a worker must possess for fulfilling a task, each characterized by the corresponding EQF level (depicting the complexity degree). Subsequently, the education domain was investigated by interviewing Education and Training Authorities of the eight countries above, with the aim of collecting information regarding learning outcomes achieved by the students at the end of a specific training route.

Results of this phase have been collected and inserted in several grids, showing information concerning units, tasks, subtasks, knowledge, skills and competences. An example is reported in Figure 3.2.

3.3.2 Taxonomy construction

In this phase, collected profiles have been inspected in order to identify core elements (knowledge objects, action verbs and context elements) to be used for constructing

Part A		**Knowledge** theoretical and/or factual knowledge		**Skills** cognitive and practical skills		**Competence** responsibility and autonomy	
List of core fields of activity and possible subtasks		Description / comments for clarification of KSC-items. Please extend the space in the cells if you need more room.					
	score		*score*		*score*		*score*
Task / Activity A: **Sells**	3,5	Knowledge of internal procedures Knowledge of basic English Knowledge of communication principles and processes	4 2,5 3	To be able to follow internal procedures To be able to understand which kind of product are more adequate to specific/special occasions	4 3,5	Responsible for the selling activity of the shop Full autonomy/with shop manager supervision	**3.5** **3**
Subtask 1: **Welcomes the customer and understands the customer's needs and requests**	4	Product Knowledge Knowledge of active listening principles and processes Knowledge of selling principles and processes Knowledge of *Ticket Pro Medium* concept	3,5 3,5 3,5 2,5	To be able to approach the customer in the appropriate timing To be able to identify and to differentiate habitual customers To be able to communicate in English with the customer To be able to understand which kind of product are more adequate to specific/special occasions To be able to apply selling techniques	3,5 3,5 3 3,5 3,5	Responsible for an effective and appropriate welcoming, for the identification of customer requests/needs and for identifying how many products can that specific customer buy Full autonomy/with shop manager supervision	**4** **3,5**

Figure 3.2. Portion of the grid of the Shop Assistant (Portugal)

the taxonomy: in particular, each instance of knowledge, skill and competence elements have been expressed as a combination of one or more concepts (or keywords).

Subsequently, identified concepts were linked to each other by subsumption relations, in a taxonomic representation, composed of three families of terms hierarchically structured: knowledge objects, action verbs and context (Figure 3.3).

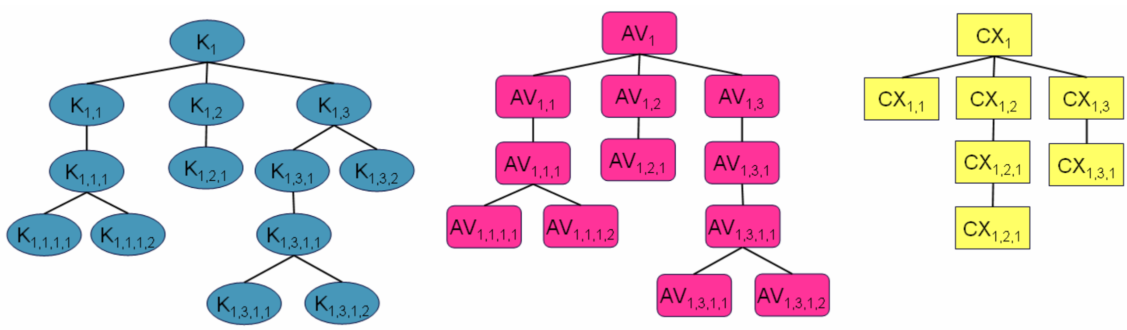

Figure 3.3. Taxonomy and families of concepts: knowledge, action verbs and context elements

For the creation of the knowledge and context trees, it was necessary to start from scratch, since the existing taxonomies were not able to fully satisfy our requirements. On the contrary, for the representation of action verbs family, an adaptation of the Bloom's taxonomy composed of six families of verbs (arrange, act, prepare, check, assess and react) has been exploited.

It is worth remarking that in the definition of the taxonomy, experts from the trade sector have been highly involved, since an improper hierarchy of concepts could provide incorrect results.

Then, after the creation of the taxonomy, qualifications, tasks and subtasks were described by linking their composing elements (knowledge, skills and competences) to the corresponding concepts (knowledge objects, action verbs and context elements), as in Figure 3.4.

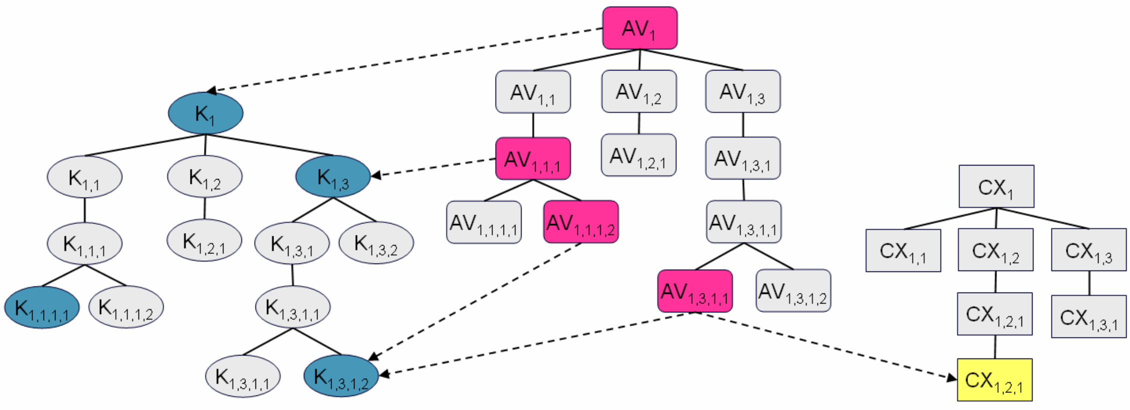

Figure 3.4. Relations among concepts

For this purpose, in order to provide a formal and easy-to-read description to be shared with the involved actors and stakeholders, a graphical representation of the relations among learning outcomes and annotated concepts was drawn by exploiting the UML notation and the open-source tool UMLGraph[1], a software that is able to process diagrams expressed in a textual form and to draw the corresponding graphical representation. UML diagrams have been embedded in the platform so as to allow users to browse work or education related maps during the population of the knowledge base.

Figure 3.5 shows an excerpt of the subtask *To welcome the customer and understand the customer's needs and requests*, belonging to the Portuguese Shop Assistant profile: in particular, the diagram displays the knowledge *Communication techniques knowledge*, the two skills *To be able to apply selling techniques* and *To be able to communicate in English*, and the competence *Full responsibility in identifying the customer and his needs.* In order to better characterize knowledge, skill and competence elements, the corresponding classes are shown in white, whereas the concepts of the taxonomy they are linked to are painted on a darker shade of color; other

[1]http://www.umlgraph.org/

knowledge objects are painted light blue, action verbs light rose and context elements light yellow. In addition, subsumption relations are expressed by a solid line with a hollow arrowhead pointing from the class that is subsumed to the class that subsumes. Finally, the fact that a knowledge, skill or competence is characterized by one or more concepts belonging to the taxonomy is denoted by a dashed line. It is worth noting that this type of lines has been used to make more readable the diagram, so that it is immediately understandable which are the relations defining subsumption of terms, and which are the relations linking knowledge, skills and competences to the taxonomy (in this case a dashed line is drawn in order to show the link between an element of a subtask and a concept in the taxonomy, or among knowledge elements, action verbs and context, and not for indicating dependency relations).

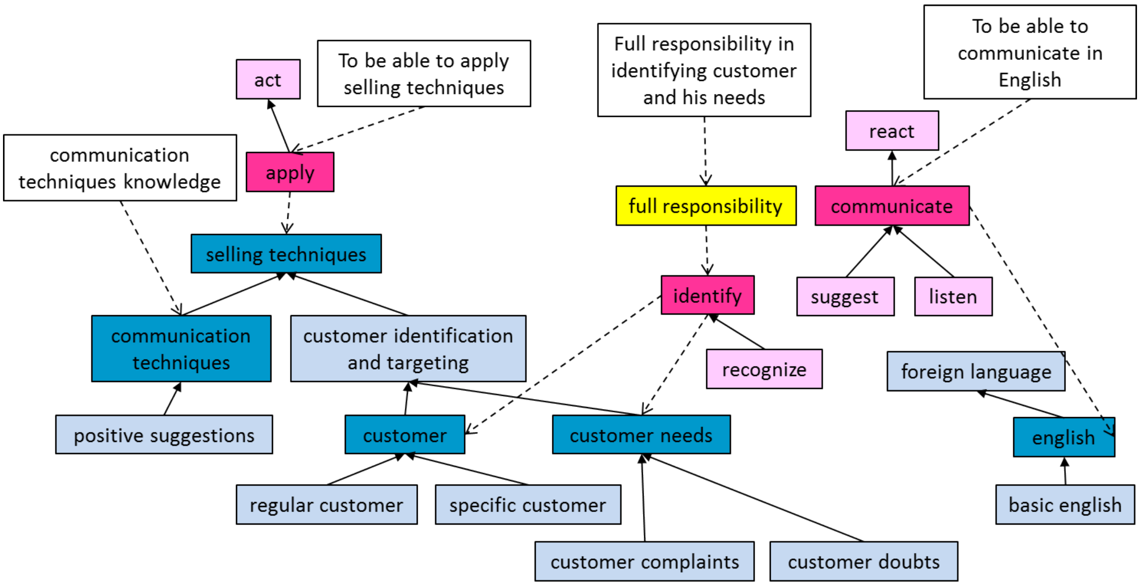

Figure 3.5. Excerpt of the graphical representation related to the subtask *To welcome the customer and understand the customer's needs and requests*

The diagram should be read as follows: the knowledge *Communication techniques knowledge* is linked to the knowledge object *communication techniques* that is a type of *selling techniques*, i.e. another knowledge object. The subsumption relation between *selling techniques* and *communication techniques* shows that if someone has got a *communication techniques* knowledge, he or she also knows something about *selling techniques*. Furthermore, the skill *To be able to apply selling techniques* is characterized by the pair of concepts *apply*, an action verb that specifies the action verb *act*, and *selling techniques*, a knowledge object, while the skill *To be able to communicate in English* is defined by the action verb *communicate*, a specification of the action verb *react*, that is applied to the *English* concept, a specification of a generic *foreign language* knowledge object. Finally, the competence *Full responsibility in identifying the customer and his needs* is described by a *full responsibility* context, applied to the *identify* action verb, that is linked to *customer* and *customer needs* knowledge objects.

3.3.3 Inference rules and approaches for semantic comparison

The basic idea behind the definition of the inference rules required for the TIPTOE project is the following: since the common profile must act as common denominator, it should be a combination of elements that are somehow expressed in all the profiles and, as a consequence, it should be the sum of all the knowledge, skills and competences that are linked to the most used knowledge objects, action verbs or context elements.

An example could probably explain in a clearer way the above statement: let us consider four subtasks belonging to four profiles, defined by the following knowledge: *cleaning techniques knowledge*, *cleaning means and tools knowledge*, *cleaning methods knowledge* and *cleaning methods, means and tools*; since each profile contains (at least) a knowledge that is related to the cleaning activity, this knowledge should also be included in the common profile. On the contrary, if only one profile mentions a knowledge, e.g. the *product lifecycle knowledge*, this knowledge shall not be incorporated into the common profile.

In addition, the reasoning that has been presented above should be based on the concepts linked to the elements belonging to the profile descriptions. In fact, a semantic engine should be able to understand that the four knowledge elements mentioned above are linked to the cleaning concept (then, they will be characterized by the *cleaning* knowledge element).

It is clear that the common profile will then be a representation of the most common knowledge, skill and competence elements. Consequently, the engine for semantic comparison should be able to identify the most used concepts, recognize which elements they are linked to, and then include these elements into the common profile. A further step towards the achievement of a more correct result could be the exploitation of the taxonomy of terms and subsumption relations: in this way, by analysing the example shown in Figure 3.5, the number of occurrences of *communication techniques*, *customer*, *customer needs*, *English*, *apply*, *identify*, *communicate* and *full responsibility* would be 1, while the number of occurrences of the (parent) element *selling techniques* would be 4, since the *selling techniques* concept has been exploited once, but the (children) elements *communication techniques*, *customer* and *customer needs* have been used each one once too.

Four comparison strategies have been developed and investigated in order to find the best result; all of them take as input a threshold (that is a minimum number of times a concept has to be used) defined by the user, and explore the taxonomy in order to identify the most common elements. The four comparison strategies developed are: *simple range*, *simple range with mean*, *aggregate simple range* and *aggregate range with mean.*

- The simplest way of determining which knowledge, skill and competence elements will belong to the common profile is the *simple range* strategy, since it calculates a value that corresponds to the number of times a concept has been linked to the learning outcomes; if this value is higher than the threshold defined by the user, the strategy includes the considered knowledge, skill or competence into the common profile.

- A slightly more complex approach is the *simple range with mean*: according to this strategy, the value computed by the comparison tool (which, in order to add the element belonging to the common profile, must be higher than the threshold defined by the user) is the average of the number of occurrences of each concept linked to the knowledge, skill or competence being considered.

- A third approach, which takes into account also hierarchical relations expressed by the taxonomy, is the *aggregate simple range*: according to this strategy, the tool calculates the number of times a concept, and the subsumed ones, have been used to describe the elements of the ontology; if this value is higher than the threshold specified by the user, the examined element is added to the common profile.

- A fourth strategy, that is similar to the *simple range with mean* and that allows to consider also subsumption, is the *aggregate range with mean*: according to this approach, the value computed by the comparison tool is the mean of the number of occurrences of each concept and its children in the taxonomy.

A further example could help in understanding the logic behind the four different approaches: let us consider the *Knowledge of products and relevant display techniques (i.e. volume displays and on shelf couponing)* element and let us assume that this knowledge is described by the concept *product* (used 38 times in the profile descriptions), *exposition techniques* (used 12 times), *volume displays* (used 3 times) and *on shelf couponing* (used only in this description). Furthermore, let us suppose that the *product* and the *exposition* elements have several children in the knowledge taxonomy, and that the respectively subsumed classes have been used 84 times and 14 times, respectively.

If the *simple range* strategy is adopted, the result is 38, that is the maximum value of occurrences of the concepts linked to the knowledge. On the other hand, the result of the *simple range with mean* approach is 13.5, that is the average of the occurrences of the four concepts linked to the element. When subsumption relations are considered, the computed value increases: in fact, the result of the *aggregate range* approach is 122, that is the sum of the occurrences of *product* (122, that is 38+84), *exposition* (26, that is 12+14), *volume displays* (3) and *on shelf couponing* (1) concepts, whereas if the strategy applied is the *aggregate range with mean*, the result is 38, that is the average of the values above.

Strategy	**Results**
Simple range	38
Simple range with mean	13.5 = (38 + 12 + 3 + 1) / 4
Aggregate range	122 = (38 + 84) + (12 + 14) + 3 + 1
Aggregate range with mean	38 = [(38 + 84) + (12 + 14) + 3 + 1] / 4

Table 3.1. Results obtained from the application of the four comparison strategies to the *Knowledge of products and relevant display techniques (i.e. volume displays and on shelf couponing)* element

Having considered all of the above, it is possible to say that, in general, "simple" approaches (*simple range* and *aggregate range*) perform better with long sentences characterized by the description of a knowledge and several examples, like *Knowledge of products and relevant display techniques (i.e. volume displays and on shelf couponing)*; but they provide worst results with skills and competences defined by a common verb and an uncommon noun (an example could be the skill *to apply stocktaking procedures*, since this element obtains a high rate, even if the *stocktaking procedure* knowledge is a rare concept). On the contrary, a strategy that computes the average of the occurrences produces better results with elements like the skill above, but risks to provide worst results in case of long sentences.

For what it concerns the exploitation of subsumption relations, if an approach that does not consider the taxonomy (*simple range* and *simple range with mean*) is pursued, all the concepts (children and parents) have the same importance but, if the different profiles are described with a huge variety of terms, incorrect results could be achieved; a strategy that considers also the taxonomy (*aggregate range* and *aggregate range with mean*) implicitly interprets as more important the highest elements of the tree (the parents) and, for this, it could be useful to overcome lexical differences.

It is worth remarking that the results just explained (and shown in Table 3.1) represent only an estimate of how common a knowledge, skill or competence is; hence, a given value could not be good or worst a priori, since it has to be compared with the other results. Consequently, possible ways for the identification of the common profile could be to order the results from the one that obtained the highest value to the one that got the lowest one, and then select a number of elements defined by the user (i.e., the number of knowledge, skill and competence elements in the common profile would be fixed), or – and this was the case for the project – to use the threshold expressed by the user to select only those elements that achieved a score higher than it.

3.3.4 Creation of the common profile

The common profile has been created by exploiting the approach explained in the previous stage. According to the above discussion, the knowledge, skill and competence elements obtaining a specific value become potential components of the common profile. However, since – like in the case of the above example with the set of knowledge described by the cleaning techniques concept – it would be redundant inserting into the common profile four elements with the same meaning, it has been decided to let the user choose, among the set of elements exploiting the same concepts, the one that could better represent the specific knowledge, skill or competence.

Figure 3.6 shows an example of knowledge described by the knowledge object *cleaning techniques*: here the user selected as the representative knowledge the element *cleaning methods*. As shown in Figure 3.7, statistics on the exploitation of single or aggregate concepts (for a single country or for the whole working or training dimension) may be exploited in this step. Also, UML maps can be navigated to get insights on specific portions of national maps regarding the education and labor domain as well as their relations with the overall profile being created (as already shown in Figure 3.5).

For the common profile, an EQF level is computed as an average of the EQF values assigned to each knowledge, skill or competence selected as potential components of the whole profile.

It is worth mentioning that, in order to perform a quick analysis of the inserted profile, the Wordle tool[2] has been exploited for the creation of a tag cloud of the composing learning outcomes.

However, even though this tool allows users to identify a profile by giving a quick look to the tag cloud (since the tag cloud of a managerial job positions will be a combination of concepts denoting higher responsibility, while an assistant job visual description will be characterized by more lower-level activities), the size of a word is only linked to the number of its occurrences in a sentence, without taking into account also relations among terms. An example is shown in Figure 3.8: in this representation, terms like *goods*, *product*, *(product) characteristics* are not related

[2]http://www.wordle.net/

to each other, while, in reality they refer to the same thing (the *product*); as a consequence, the number of occurrences of this group of terms, instead of the value expressing single occurrences, should be considered.

Figure 3.6. Selection of the elements that will belong to the common profile

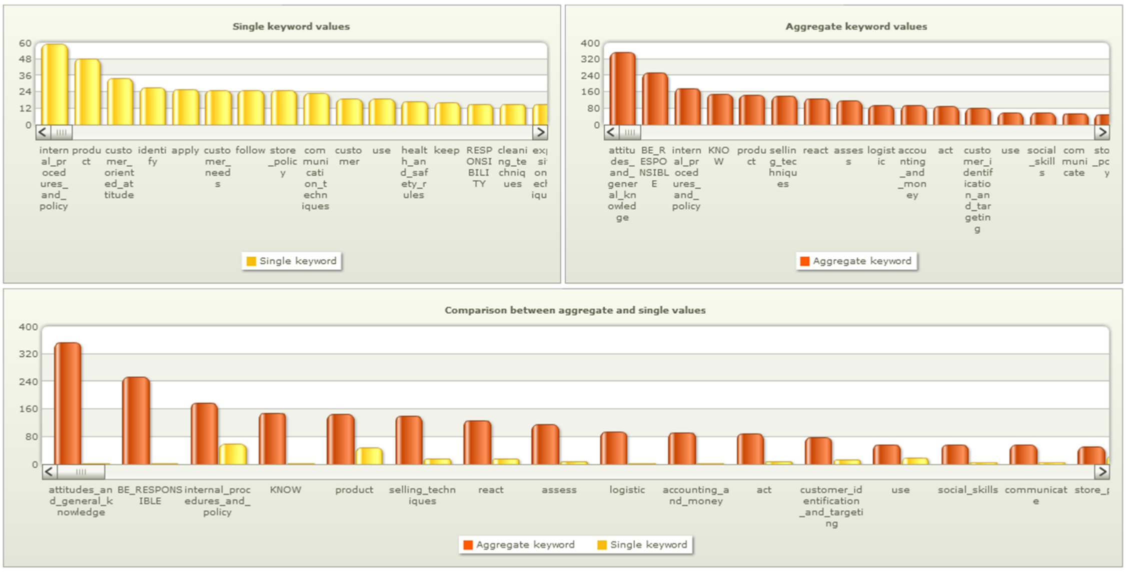

Figure 3.7. Statistics on the exploitation of concepts in the description of learning outcomes characterizing the Shop Assistant job profile

3.4 Services targeted to end-users

In addition to the features described in the previous Section, a tool for the automatic identification of the EQF level of a given qualification based on information stored in the knowledge base and the *EQF ruler*, a navigable collection of shared learning outcomes for the trade sector structured according to their EQF level, have been implemented.

Figure 3.8. Tag cloud of the learning outcomes related to the Shop Assistant profile

3.4.1 Automatic identification of the EQF level

The tool for the automatic identification of the EQF level allows users to insert in the platform a new profile, and to exploit the semantic engine (used for the construction of the common profile) for the automatic identification of the corresponding EQF level. In this way, an immediate comparison of the owned knowledge, skills and competences with respect to the European reference can be obtained. The underlying idea behind the development of this instrument is that for each learning outcome belonging to a given profile, an EQF level could be automatically identified by considering the EQF level previously assigned to similar learning outcomes.

For evaluating the applicability of such a tool, the referencing of a profile for international trade developed in the framework of the COMINTER project[3] has been considered as a test bed. First, the core elements (units, task, subtasks and learning outcomes) belonging to the COMINTER profile have been inserted into the platform together with a set of concepts (i.e. knowledge object KO, action verb AV and context elements CX).

Afterwards, starting from the information collected in the knowledge base, the suite for the identification of the EQF level has been created. The recognition of the correct level is carried out by performing a semantic search on concepts expressing the meaning of the unlabeled learning outcome, in order to identify whether the knowledge base yet contains some knowledge, skills or competence (with an EQF level previously specified) exploiting the same concepts used for the description of the COMINTER element. However, since a perfect match between the searched learning outcome and the already inserted knowledge, skills and competence is nearly rare (because both elements should have been described by the exact set of concepts), it

[3]http://www.cominter-europe.org/

is evident that a search engine should also browse the taxonomy (in order to identify more specific and more general concepts) and that scenarios in which only a part of the whole set of terms is contained in the semantic description of a learning outcome must be considered.

An example could better clarify the above statement: let us consider a skill *To be able to recognize the needs of a customer* whose EQF level is unknown and let us imagine that the linked concepts expressing its meaning are *recognize* and *customer needs*. The search engine should then look inside the knowledge base in order to identify all the elements described by that exact set of terms. Even though these concepts are frequently used, let us suppose that no other learning outcomes are described by the couple *recognize* - *customer needs*, but there are some skills expressing the ability of identifying customer needs (thus, described by the couple *identify* - *customer needs*), and that the underlying taxonomy of terms somehow defined a relation between the *recognize* (more specific) and the *identify* (more general) concepts, as already depicted in Figure 3.5. In this case the former skill is a specification of the latter skill, since, according to the taxonomy, the activity of recognizing customer needs implicitly requires the ability to identify customer needs. Consequently, the two skills could be considered as similar, and the EQF value of the latter could be automatically assigned to the former.

In order to achieve the aforementioned result, the tool for the identification of the EQF level first identifies all the possible combinations among the sets of concepts used for the description of a given learning outcome (i.e., KO, AV and CX) by also browsing the taxonomy; then, it performs a search in the knowledge base to identify the presence of already inserted knowledge, skills and competences described by one (or more) combinations of terms (in this example, possible combinations are *recognize* - *customer needs*, *recognize* - *customer identification*, *recognize* - *selling techniques*, *identify* - *customer needs*, *identify* - *customer identification*, *identify* - *selling techniques*). It is worth remarking that, in this phase, different knowledge, skills or competences already present in the knowledge base could be identified, according to the different combinations of terms (for example, several learning outcomes containing the couples *recognize* - *customer needs* or *identify* - *customer needs* could be found): in this case the tool privileges the ones described by the more specific concepts (i.e. *recognize* - *customer needs*), since they are closer to the meaning of the unlabeled element.

Once the reference learning outcomes have been picked out, the EQF level is computed as an average of the values characterizing them. However, together with this value, additional information concerning its reliability should be displayed to the user. Hence, a measure of the distance between the searched concepts and the concepts found in the taxonomy is reported: this value gives an idea of the remoteness of two elements and it is minimal in the case of coincident concepts. Thus, for two couples of concepts *recognize* - *customer needs* and *identify* - *customer* *needs* the distance will be equal to one, since, in the taxonomy, the *identify* AV is only one level higher than the *recognize* AV.

Another possible scenario the user should be aware of occurs when only a limited number of concepts describing an element of the COMINTER profile is contained in the description of a knowledge, skill or competence. Let us clarify the above statement with the additional example of an unlabeled skill *to be able to recognize the needs and the expectations of a customer.* In this case the meaning of the learning outcome is expressed by the set *recognize customer needs* and *customer expectations*. Let us compare this skill with an already inserted one, like *to be able to identify customer needs*, and let us imagine that this is the closest element in the knowledge base (hence, no other learning outcomes described by the three concepts or by a combination of their higher level elements could be identified): in this case, even though skills denote similar things, it is evident that the first is more complex than the second one, since it requires also the ability to recognize customer expectations. This peculiarity should be communicated to the user, in order to make him aware of the accuracy of the EQF value that has been identified. Hence, together with the result, a value reporting the percentage of concepts found in the semantic description of a learning outcome is also shown.

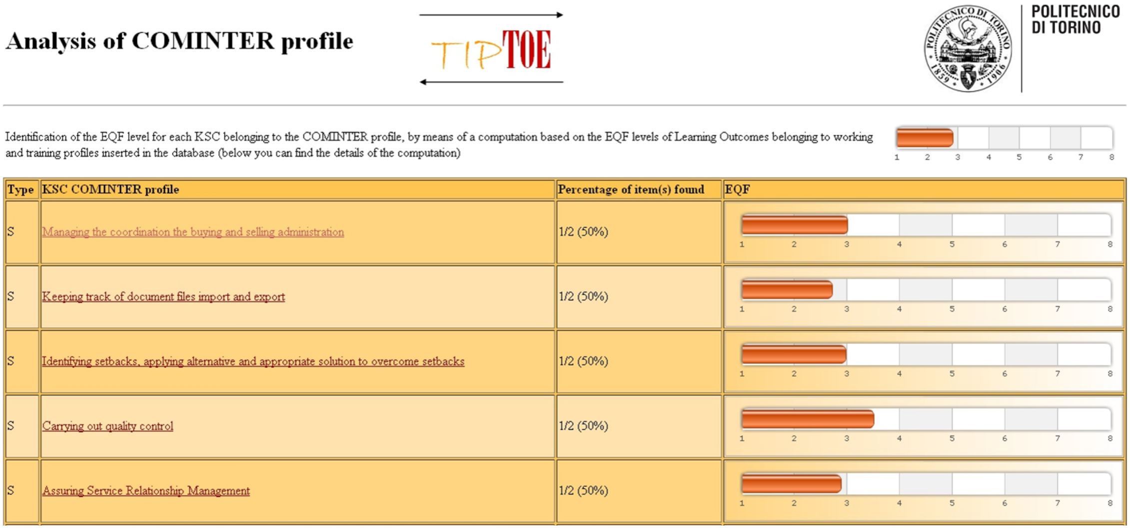

Figure 3.9. Results provided by the tool for the identification of the EQF level

Figure 3.9 shows a snapshot of the page computing the EQF level for different aggregations (learning outcomes, units, etc.). As a matter of example, the EQF value assigned to the considered unit is slightly less than 3, since it is the average of the levels of the composing knowledge, skills and competences.

It is worth remarking that, while, on the one hand, end-users could benefit from such an instrument, on the other hand, each time they insert a new profile for determining its EQF level, they contribute to broaden the knowledge base itself. Hence, a massive exploitation of the TIPTOE web portal by users and stakeholders simultaneously contributes at enriching contents and improving results.

3.4.2 The EQF ruler

The *EQF ruler* is a shared collection of learning outcomes for the trade sector structured according to their EQF level and to the type of the task they are referred to, which has been developed in order to provide stakeholders with an easy way to access the core outcome of the project, i.e. the European profile.

For what it concerns occupations related to the retail dimension, five task areas have been identified: *sales and customer relations*, *goods processing*, *presentation, promotion and marketing*, *money* and *shop management*. Figure 3.10 shows an excerpt of the *EQF ruler* created for Shop Assistant and Shop Manager occupations: the first column displays a synthesis of the required degree of autonomy and responsibility, whereas columns from the second to the fifth provide information on the context and possible tasks linked to the task areas. It is worth noting that in this case the *shop management* area has not been displayed, since EQF levels 2 and 3 do not require any management activity.

	Description of the occupational level	Task areas			
		Sales and Customer relations	Goods processing	Presentation, promotion and marketing	Money
EQF 2	Acting, carrying out activities (of tasks) Responsibility: basic, no responsibility for others Authonomy: some/none (only for routine decisions and selfmanagement), under direct supervision Adaptation behaviour to: • structured context, applies health and safety regulations • routine problems • predictable changes	Context: • Average customers • Answering simple questions of customers, referring to colleagues • Working under direct supervision Possible tasks: • Greets the customer (according to store policy)	Context: • Clear work instructions and procedures • Working in the right velocity and with good quality • Working under direct supervision Possible tasks: • Receives goods, unloads and unpacks goods, checks goods for quantity and quality and stores goods (under proper conditions) • Helps with the stock inventory	Context: • Clear work instructions and procedures • Working in the right velocity and with good quality • Working under direct supervision Possible tasks: • Prepares products for sale, applies price tags and labels to products • Restocks shelves, checks quality of products on shelves and arranges products on shelves in a neat way	Context: • No activities with respect to payments and cashiers • Only assisting activities Possible tasks: • Prepares and packs the purchases according to customer's whishes
EQF 3	Acting, carrying out tasks Responsibility: some responsibility, responsible for own work, no responsibility for others, reports to his/her manager on his own activities, sales results and other issues regarding shop performance, responsible for reaching own targets Authonomy: some/low, following procedures for solving problems, problem solving in low complex situations, under direct and indirect supervision Adapting behaviour to: • structured context, familiair circumstances • routine problems • predictable changes Able to react properly in new situations	Context: • Average customers • Answering questions of customers in a professional way, giving information in a friendly way and selling simple products • No financial targets for selling Possible tasks: • Greets and/or approaches the customer (according to store policy) • Determines the customer's needs and wishes • Advises and informs the customer using: active listening techniques, basic communication and selling techniques and basic product knowledge • Demonstrates products, informs customers about product characteristics • Suggests additional service/products to the customer • Handles returns and reclamations according to procedure	Context: • Clear work instructions and procedures • Working in the right velocity and with good quality Possible tasks: • Receives goods, unloads and unpacks goods, checks goods for quantity and quality, stores goods (under proper conditions) and administration of the activities • Makes stock inventories • Orders merchandise (following procedures) • Handles products under the right conditions • Notices logistic needs and suggests improvements	Context: • Clear work instructions and procedures • Working in the right velocity and with good quality Possible tasks: • Prepares products for sale, applies price tags and labels to products • Restocks shelves, checks quality of products on shelves and arranges products on shelves in a neat way • Creates (special) product presentations and displays, decorates the shop window • Assists management with promotion activities, makes suggestions about sales promotions	Context: • Clear work instructions and procedures • Accurate working • Handles only regular, routine cash, payments and return transactions Possible tasks: • Makes the till ready for use, attends and closes the till • Charges products • Handles regular cash and other payment means • Prepares and packs the purchases according to customer's whishes • Handles regular exchange and return transactions

Figure 3.10. *EQF ruler*: occupations for Retail

The *EQF ruler* additionally proved to be a good way to help stakeholders familiarize with the EQF: in fact, the framework is often seen as a theoretical instrument, and involved parties find it extremely difficult to exploit it practically (e.g. for recruitment, job-seeking, etc.). A survey confirmed that this instrument was fully appreciated as a tool that simplifies the whole referencing process since it tells the user the work context of an individual per level: hence, it could be considered as a valid starting point for professional users as it allows them to discuss EQF levels issues by making them more visible and easily perceived.

3.5 Conclusions

As presented in this Chapter, the research activities carried out within the TIPTOE project, were aimed at defining a semantic-based methodology for the construction of a common EQF-aware European profile in the trade sector merging contributions from national educational and occupational profiles. The objective of the presented web platform was twofold.

On the one hand it was aimed at supporting involved actors in the creation of a common profile for qualifications related to the European trade sector. On the other hand, it provided end-users with a set of services supporting them in the analysis of qualifications expressed according to the EQF principles. The common profile created within the project could lower barriers linked to information asymmetries between European labor and training dimensions, as it defined a common basis for the construction of qualifications belonging to the European trade sector. Moreover, the suite of services for the automatic identification of the EQF level could support end-users in the analysis of a new profile, and could better assure transparency during qualification recognition phases. Additionally, the *EQF ruler* could be used by them during interviews with job candidates in order to assign an EQF level to learning outcomes characterizing job applicants' grounding, thus making them fully understand whether the person in front of them is too much or not enough skilled, with respect to a searched job profile.

However, even though the proposed platform was able to reduce the workload for end-users, during the creation of the common profile, it required a huge effort in the creation of the taxonomy and in the annotation of inserted profiles. Consequently, the second part of the Ph.D. research activities have been strongly devoted to the identification of possible solutions to overcome this limitation (e.g. exploitation of already existing taxonomies, development of tools for automatic annotation, etc.).

These solutions are proposed in Chapter 4, where a platform exploiting an already existing semantic thesaurus for annotating learning outcomes is presented.

Chapter 4

Understanding the semantics of job offers and demands in a job matchmaking scenario

4.1 Introduction

In the last decade, validation and capitalization of formal, non-formal and informal learning in the perspective of citizens' mobility and employability became a key issue in the European legislation. Several initiatives were launched to support the development of suitable strategies for improving both the European education and training as well as the labor market areas, by re-defining the learning outcomes expected from existing learning paths and re-designing related supporting instruments in a way that they could allow students and workers to personalize and complement their training with the aim of seeing their competences recognized both in the school and the labor worlds. One of the obstacles to the achievement of these objectives has traditionally been represented by the lexical and semantic differences between the descriptions of education syllabi, personal achievements, expected abilities and so forth. Important steps to address the above constraints were done with the definition of tools like the Europass portfolio[1], the European Qualification Framework, the European Credit Transfer System (ECTS)[2], etc. The goal of the above tools was to improve readability and transparency of learning outcomes and individual skills in a European-wide perspective. Such tools were meant to support schools (looking for a way to compare qualifications), students (looking for education and training paths capable of filling their learning gaps), workers (looking for job positions where

[1] http://europass.cedefop.europa.eu/en/documents/curriculum-vitae

[2] http://ec.europa.eu/education/lifelong-learning-policy/ects_en.htm

their abilities could be best valorised), companies (looking for the best people with the right competences to hire), etc.

Despite the key role that is expected to be played by the above tools in the European dimension, an important issue that has still to be explored is their applicability in contexts where qualifications and skills owned by migrants (i.e., people from non-European countries) have to be considered. In this case, equivalence rules enabling a comprehensive comparison of qualifications from different education and training systems are often unavailable. Moreover, there are situations (like, for instance, in the human resource acquisition phases) where a check on owned qualifications is not sufficient. In fact, in these cases, job-seekers' competences possibly achieved in non- formal contexts have to be analytically matched against skills needed for the particular job offer. Information asymmetries in the above situations may threaten the competitiveness of the education and labor worlds. Hence, tools capable of supporting the matchmaking (i.e. the process of matching offer and demand) between job seekers' skills, companies' requirements and education profiles by working on detailed descriptions of qualifications, résumés and labor market's needs are required. These instruments should be capable of comparing the above descriptions based on their inner structure and contents. In this light, the exploitation of semantic tools, could increase the effectiveness of matchmaking, since they could allow a computer system to understand and to (automatically) process the huge amount of heterogeneous data and relations involved in the analysis tasks.

Based on the such considerations, activities carried out during the Ph.D. have been addressed to tackling heterogeneity issues in the descriptions of qualifications, résumés and labor market's needs due to the use of non-shared vocabularies. Research has been performed in the context of the MATCH "*Informal and non-formal competences matching device for migrants' employability and active citizenship*" project.

The aim of the MATCH project was the development of a web-based functional tool connecting the migrants'/job seekers' competences acquired in formal, non-formal and informal contexts to occupational profiles and to companies' labour demand. The devised platform, the LO-MATCH platform, aimed at supporting Chambers of Commerce and training organizations in the effective link of employers' needs with job seekers characteristics in order to find the best, or at least good, solutions for the involved parties (thus, by solving a matchmaking problem). Such a tool had to be able to:

- allow migrants/job seekers to identify, among the job offers contained in the knowledge base, the ones that could better valorize their competences. This could be done by comparing migrants'/job seekers' characteristics with companies' requirements;
- help companies to select best candidates for a job position, by providing them

with a ranked list of job applicants that show competences required for performing the specific working activity;

- support migrants/job seekers in the identification of competences they should acquire in order to increase their possibility to be employed by a given company.

As in Chapter 3, the main issue was to find commonalities and differences within data (in this case, curricula and job offers) expressed with different formats and at different levels of detail. However, in this Chapter, rather than the exploitation of an ad-hoc taxonomy, an already existing semantic thesaurus, WordNet, has been used to deal with contents from different countries and related to different contexts (e.g. from vocational to higher education, from the mechanics to the construction sector, etc.).

The rest of Chapter is organized as follows: Section 4.2 describes the reasons behind the adoption of a semantic-based approach, together with the characteristics of the WordNet semantic thesaurus; Section 4.3 presents the four phases of the creation of the LO-MATCH platform, whereas the overall architecture is drafted in Section 4.4. Then, functionalities and additional features are shown in Section 4.5 and Section 4.6 respectively. Finally, conclusions are drawn in Section 4.7.

4.2 The need for a common language: semantically describing curriculum vitae and job offers

The goal of the MATCH project was to make it possible that when a migrant/job seeker draws his or her curriculum vitae, this is expressed in a way that is "compliant" to the way employers/recruiters have expressed their requirements (and vice versa), so that the best match between job offer and demand could be actually found, independent of possibly existing barriers to a shared understanding. To maximize the overlap, information pieces involved in the overall matchmaking process depicted above had to be expressed by means of a common language. To deal with such issue, like various works reported in the literature and commercial platforms for e-job seeking and recruiting available online, the MATCH project chose to pursue a semantic-based approach. Thus, the basic idea was to create a vocabulary shared among all the actors involved to be used for describing both the acquirements in job seekers' profiles as well as the requirements in employers' job offers.

As a matter of example, with non-semantic techniques, information like *to work out design sketches by hand*, *to use under supervision CAD applications for the fashion industry*, *developing technical drawings using a PC program*, *to autonomously* *exploit current apparel software applications*, and *to draw dresses manually* are generally compared by using a keyword-based approach. Thus, some kind of match among them could only be found where the common terms *application* and *draw*, in this case, are used. On the contrary, in a semantic-aware vision, raw keywords used are associated with concepts, which are assigned a particular meaning and linked to other concepts by means of various kinds of relations. This way, concepts are organized in taxonomical and/or ontological knowledge models, which can be leveraged to develop automatic reasoning algorithms able to navigate such structures are apply inference rules to deal with information asymmetries and solve the matchmaking problem addressed. For instance, the relation between *to draw*, *drawings* and *sketches*, or between *CAD*, *PC program* and *software* would be made explicit (Figure 4.1).

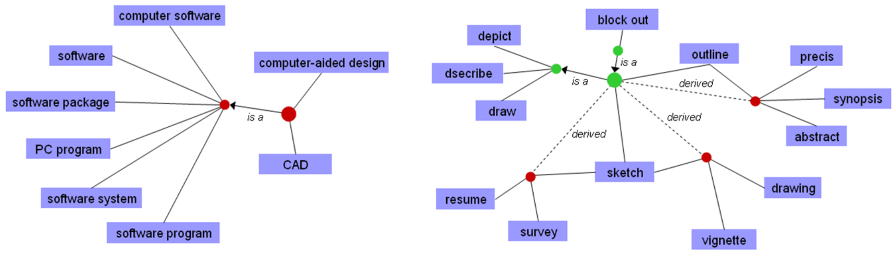

Figure 4.1. Relation between *CAD*, *PC program* and *software* or between *to draw*, *drawings* and *sketches*

According to the literature presented in Chapter 2, the construction and maintenance of such formalizations are extremely time consuming. In fact, they require the intervention of so called "knowledge engineers", i.e., experts of both the domain of interest as well as of the sophisticated tools, languages, etc. associated with the semantic technology. Moreover, it has been showed that the effectiveness of the overall semantic-based processing is strictly related to the breadth and quality of the models developed, which also need to be kept up-to-date in order to maintain their applicability (for instance, in the example above, all the possible kinds of software programs should be considered, and updated over time e.g. taking into account advancements in the field). Also, in many cases there is ambiguity or confusion in the way a given concept has to be semantically interpreted. For instance, in many works, job seekers' curriculum vitae and job offers are regarded as containers or competency elements, in other cases of skill items. And, as said, the way competency, skill and other concepts are intended could change a lot based, among others, on the geographical, systemic or sectoral context. Finally, in some situations it has also been shown that creating strict (generally hierarchical) structures, e.g., taxonomic or ontological, between concepts like knowledge, skills and competences could be even counterproductive when some of the elements that could contribute to their definition like the proficiency level, or the context, etc. cannot be taken into account. To make an example, what is the actual relation between the *programming* and the *debugging* ability? Is it always correct assuming that if an individual is able *to write a software program*, then he or she is also able *to debug it* (or vice versa)?

The approach pursued in MATCH has been designed by taking all the considerations and critical points above into account. First, the central harmonization role of the learning outcome concept has been considered. That is, job seekers' acquirements and employers' requirements are expressed by making reference to knowledge, skill and competence elements as defined in the EQF. This is a novelty for the literature on (semantic) job matchmaking, mainly because most of the software solutions available have been presented before the establishment of the EQF itself or because they have been developed by neglecting the lifelong learning constraints (at least the European ones). Afterwards, in order to alleviate the burden associated with the management, in the broad sense of the term, of shared formal models, the construction of the knowledge base started with the collection of educational and occupational profiles. No knowledge of semantic-related concepts was requested at this stage by involved parties, since profiles were simply described in terms of tasks, sub-tasks and learning outcomes.

Then, the transformation of the above collection of profiles into a semantic-aware knowledge base was due to a semi-automatic pre-processing step, which exploited an existing general-purpose ontology together with the definitions given in the EQF recommendation and in the Europass guidelines (also exploited in previous works) to produce the required formal domain model.

That is, each learning outcome was first considered against the definition of a competence as the ability of putting into action a given set of knowledge objects in a specific context. Thus, action verbs, as defined in the glossary developed by Cedefop[3], were identified and linked to knowledge objects and to context elements. Still referring to the example above, based on the approach described, in the learning outcome *to autonomously exploit current apparel software applications*, the action verb *to exploit*, the knowledge object *current apparel software applications* and the context element *autonomously* would be identified.

After having performed this annotation step in an automatic way, domain experts from partner organizations were allowed to adjust results achieved by using a graphics tools embedded in the platform, which succeeded in completely hiding to unskilled operators the complexity of underlying semantics. Then, by taking into account limitations encountered during the TIPTOE project, and presented

[3]http://europass.cedefop.europa.eu/en/documents/european-skills-passport/certificate-supplement/action-verbs-glossary

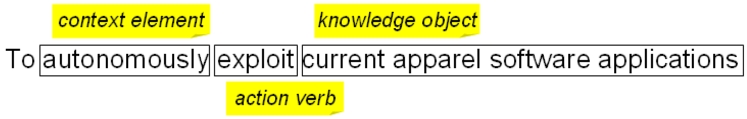

Figure 4.2. Knowledge objects, action verbs and context elements in the learning outcome *to autonomously exploit current apparel software applications*

in Chapter 3, instead of improving the model by creating taxonomies of concepts (e.g., to relate more general knowledge elements to more specific ones by means of subsumption relations, etc.), terms founds for action verbs, knowledge objects and context elements were linked to concepts defined in WordNet, and the final ontology was created by leveraging on semantic relations defined therein (to be possibly enriched by domain experts, hence the use of the "semi-automatic" definition).

WordNet is a large lexical database of English. It can be browsed online (by using the official text-like application[4] or one of the graphics tools available, like WordVis[5], Visuwords[6], etc.), or it can be downloaded and exploited as a software library for building new software tools (like in the case of MATCH)[7]. It is not specific for the particular domain considered. That is, content described is valid for many purposes (e.g., it can be applied to job matchmaking or to other application scenarios, it can be exploited to deal with various sectors, etc.) and users do not have to be experts of a given context for managing it. In WordNet words (i.e. nouns, verbs, adjectives and adverbs) collected in the WordNet database are grouped in a set of synonyms called synsets, each one expressing a distinct concept by providing a short general definition. In addition, information about semantic relations among synsets is recorded.

While, on the one hand, WordNet produces a combination of dictionary and thesaurus that is more intuitively usable, on the other, it supports automatic text analysis and artificial intelligence applications. On WordNet's website, several software tools as well as the semantic database are freely available for download. In addition, the semantic database could also be browsed online.

Although, at first sight, WordNet could resemble a thesaurus, since it groups words together based on their meanings, it has several additional features: first, it interlinks not just word forms (i.e. strings of letters) but specific senses of words; second, it also labels the semantic relations among words, whereas groups of words,

[4]http://wordnetweb.princeton.edu/perl/webwn

[5]http://wordvis.com/

[6]http://www.visuwords.com/

[7]http://wordnet.princeton.edu/wordnet/download/

in a thesaurus, are only created on the basis of meaning similarity. In the WordNet view, synonyms, the more frequent relations among words, are grouped into unordered sets called “synsets”. Each one of WordNet’s 117 000 synsets is linked to other synsets by means of a small number of semantic relations. In addition, for each synset, a brief definition is provided, together with, in most of the cases, one or more short sentences illustrating the use of the synset members. The most frequent relations among synsets are *hypernyms* (Y is a hypernym of X if every X is a (kind of) Y) and *hyponyms* (Y is a hyponym of X if every Y is a (kind of) X), that are super-subordinate relations. Moreover, WordNet makes a distinction among Types (common nouns) and Instances (specific persons, countries and geographic entities). Thus, *armchair* is a type of *chair*, *Barack Obama* is an instance of a *president*. It is worth remarking that Instances could only be leaf (terminal) nodes in their hierarchies. A less frequent, but still important relation is *meronymy* (X is a meronym of Y if X is a part/member of Y). It is worth remarking that, unlike hypernyms and hyponyms, for which relations are transitive (i.e. if an *armchair* is a kind of *chair*, and if a *chair* is a kind of *furniture*, then an *armchair* is a kind of *furniture*), parts indicated by meronymy are not directly inherited, since they may be characteristic only of specific kinds of things rather than the class as a whole (*chairs* and kinds of *chairs* have *legs*, but *not* all kinds of furniture have *legs*).

Verb synsets are arranged into hierarchies as well. In addition, relations such as *troponymy* (the presence of a “manner” relation between two synsets, like in *communicate-talk-whisper*, where the manner depends on the volume) are depicted in the database.

Adjectives are organized in terms of *antonymy.* In particular, when two adjectives are linked through an antonymy relation (like *wet-dry*), they are called “direct antonyms”, whereas, when semantically similar adjectives are considered (by comparing, as a matter of example, *arid*, linked to *dry*, with *soggy*, linked to *wet*), the antonymy relation is “indirect”. In addition, other relations, such as *pertainyms* (pointing to the nouns adjectives are derived from, like *criminal-crime*) are reported.

This way, the preliminary knowledge base made up of learning outcomes annotated according to EQF and Europass definitions was turned into an ontological model with a tight network of relations, from very straightforward to very sophisticated. That is, still referring to the example already discussed, besides being described by annotations about its constituting action verbs, knowledge objects and context elements, a learning outcome like *to autonomously exploit current apparel software applications* could be compared with *to draw dresses manually* (because of the relation between *apparel* and *dress*) as well as with *developing technical drawings using PC a program* and *to use under supervision CAD applications for the fashion industry* (because of the relations between *software*, *PC program* and *CAD*, among others).

The way comparison could be exploited in an extremely simplified matchmaking

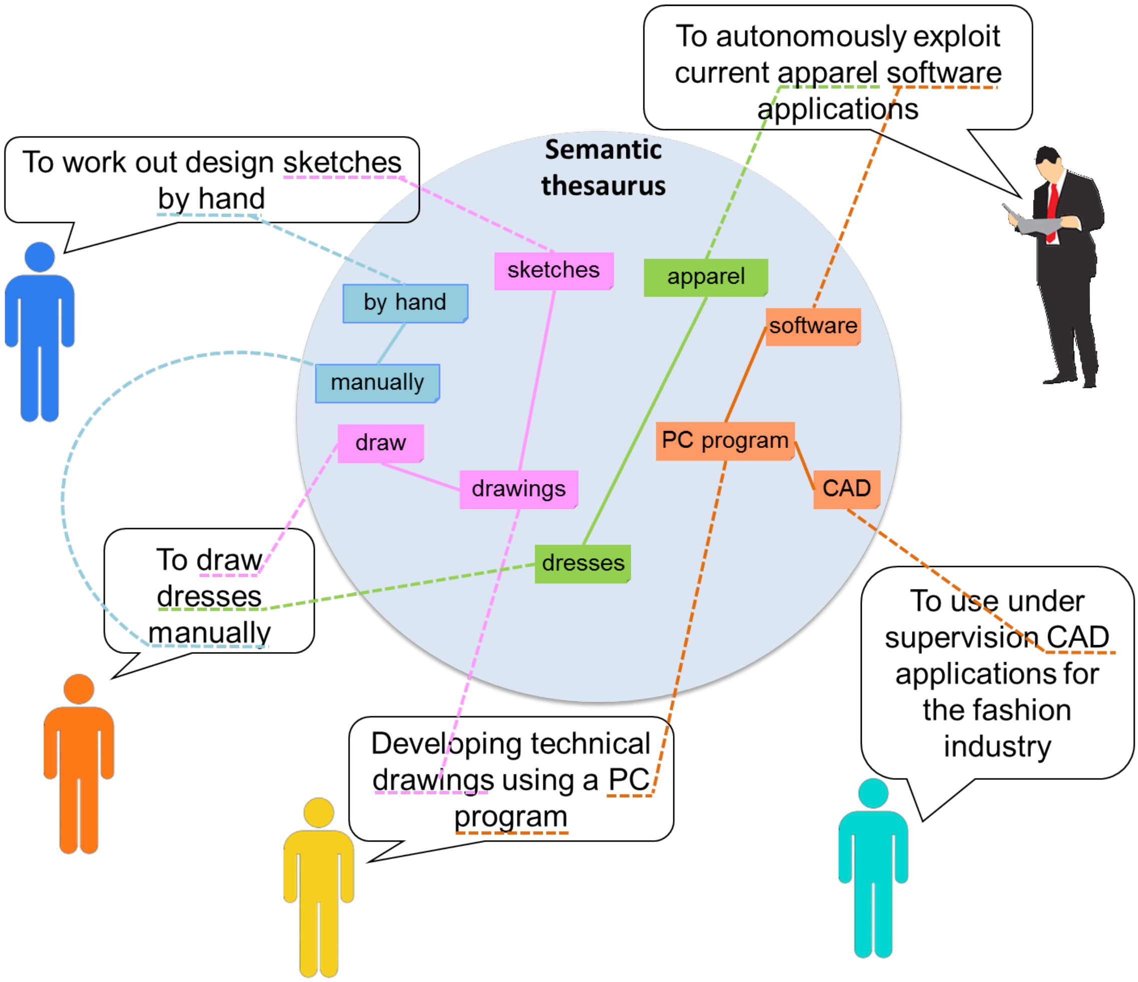

Figure 4.3. Example of comparison of annotated learning outcomes

scenario is illustrated in the Figure 4.3, where the elements available in the developed model are used to annotate and later match learning outcomes involved in the depicted job seeking - job recruiting deal (annotations are represented by coloured notes and dotted lines, the WordNet semantic thesaurus is represented by the blue sphere, whereas concepts are depicted by means of boxes and relations among them by means of solid lines among boxes).

4.3 Creation of the LO-MATCH platform

The development of the LO-MATCH platform could be summarized in four main phases: a) collection of professional figures/qualifications; b) their annotation in the knowledge base; c) population with curricula and job offers d) computation of the match.

4.3.1 Collection of professional figures/qualifications

In this phase, project partners (that are Chamber of Commerce, Training Authorities, Vocational Education and Training providers associations, etc.) investigated their country's employment situation in order to identify professional figures and qualifications to be inserted in the LO-MATCH knowledge base. This step had been carried out by having in mind the project's objective, thus, by trying to find which were the most requested professional figures, or which were the sectors that could benefit at most from the exploitation of the LO-MATCH platform. The result of this phase was a list of learning outcomes that are provided by a qualification, or a set of tasks, activities and related learning outcomes that should be possessed in order to perform a given job. This information was collected by means of desk research or interviews with relevant stakeholders.

Table 4.1 shows professional figures that have been inserted by project partners in the LO-MATCH platform.

Collected information was then structured in a form compliant with EQF guidelines (hence, in terms of learning outcomes or knowledge, skills and competences elements), by taking also into account the methodology developed in [Pernici et al. 2006] and already presented in Chapter 3.

4.3.2 Annotation of professional figures/qualifications

As quickly presented above, for the annotation of professional figures and qualifications, a specific facilitator, exploiting the WordNet semantic thesaurus and presenting concepts and relations among them in a graphical way, has been developed. This way, each time a learning outcome is inserted in the platform, the system automatically identifies concepts composing the sentence and gives to the user (the project partner) the possibility to specify by means of which terms he/she would like to annotate the learning outcome, and to select the more appropriate meaning (i.e. the WordNet synset) for the specific concepts. Hence, as a matter of example, when the user inserts the learning outcome *to use equipments for the storage of food and drinks* belonging to a *Bartender* professional figure, he/she could decide to annotate it through the *use*, *equipment*, *storage*, *food* and *drink* concepts; the platform then would ask him or her to further define selected concepts, hence, he/she could select for the *use* concept the synset *put into service – make work or employ for a particular purpose*, for *equipment* the synset *an instrumentality needed for an undertaking or to perform a service*, and so on. Moreover, he or she could also decide

Table 4.1. Professional figures inserted in the LO-MATCH platform

Italy
Bartender
Motor vehicles body repairer
Deliverer
Mechanical Maintainer
Leather tanner
Warehouse Keeper
In-home caregiver
Cleaning Assistant
Waiter
France
Mason
Cook
Spain
Health and social care to people at home
Geriatric care in social institutions for dependent people
The Netherlands
Shop Assistant
Logistics Assistant
Poland
Waiter
Cook
Fast food employee
Slovenia
Shop assistant
Cleaner's assistant

to put together one or more concepts, to write new synsets for those terms that are not contained in WordNet, and to link them to already existing synsets.

It is worth remarking that, during the annotation, if the words composing a learning outcome are not include into WordNet, the algorithm for annotating them manipulate the terms, in order to identify the root. This step is performed as follows:

- the algorithm exploits WordNet in order to identify if the word is a noun or a verb;
- if the word is a common noun, singular and plural forms are matched by removing *-s* and *-es* suffixes;

- if the word is an irregular verb, the algorithm identifies it, thanks to the morphology WordNet's view, that contains irregular forms of English words;
- if the word is a regular verb, suffixes like *-er*, *-ing*, *-ed* are removed from the term;
- in all the other cases, the algorithm removes suffixes like *-s*,*-es*,*-ed*,*-er*,*-ing*.

Moreover, terms that are not relevant for the matchmaking of job offer and demand, such as articles, conjunctions, etc. are not considered for annotation.

4.3.3 Population with curricula and job offers

In this phase, some migrant's résumés, companies requests and qualifications have been inserted in the knowledge base. In order to simplify this task, a web page similar to the one of existing search engines has been developed (views of the interface will be shown in Section 4.5.3). In this page the migrant (or company) could write owned (or searched or provided) learning outcomes, and could see a ranked list of pre-annotated learning outcomes that could be similar to the one just typed. Annotation of the inserted learning outcome would then made simple, since a migrant could select one or more learning outcomes that are similar to the one he/she wrote, and thus, he/she could exploit their linked concepts and synsets to annotate the just typed learning outcome. Hence, when a migrant inserts the skill *(to be able to) utilize aliments' warehouse*, the platform would browse WordNet and would encounter the following relations: *utilize* is synonym of *use* (they share the same synset), the synset of *aliment - a source of materials to nourish the body* - is linked to the synset of *food - any substance that can be metabolised by an animal*, whereas, similarly, the *warehouse* concept's synset is linked to the one belonging to the *store* verb. Similarly, a company could write a job advertisement containing, among other learning outcomes, the skill *(to be able to) stock beverages in the pantry*, and could automatically annotate it by finding a similar learning outcome in the knowledge base. Once annotation has been performed, demand and offer could be compared, and a matchmaking algorithm would rank job offers or demands.

4.3.4 Computation of the match

In order to rank job offers and demands, the LO-MATCH platform computes the similarity between acquired and required learning outcomes by considering composing concepts as follows: supposing that a learning outcome has been annotated through N concepts, the degree of similarity will be the result of

$$sim = \left(\sum_{i=1}^{n} \left(\frac{x_i}{N} \right) \times s_i \right) \times l \tag{4.1}$$

where:
x_i is equal to 1 if the i-th concept has been found, 0 otherwise
s_i provides information on the degree of similarity between a searched concept and a found one. It could assume values between 0 and 1, with 0 meaning concepts completely different and 1, meaning equal concepts. It is worth remarking that the user can increase/decrease the speed according to which this value decreases as the distance between concepts in the WordNet network of relations increases
l is a corrective coefficient influences by the length of the learning outcome as follows:

$$l = (1 - min) \times B^{-|d_0 - d_f|} + min \tag{4.2}$$

with:
min being the lowest value that l can assume
B representing the sensibility of the expression to the gap between distances
d_0 being the average distance between the words of the searched learning outcome
d_f being the average distance between the words of the analysed learning outcome.

By varying min and B parameters, the end-users can modify the behaviour of the function for computing the match as follows: lower min will penalize long learning outcomes, whereas, lower B will result in a corrective coefficient l slowly changing.

4.4 Overall architecture of LO-MATCH

After having discussed from a rather abstract point of view the process that led to the construction of the knowledge base underlying the LO-MATCH platform, the actual software architecture that supported its implementation will be reviewed. As it can be seen from Figure 4.4, software and hardware design followed a modular strategy that is based on incremental steps and, for this, it is quite close to the approach that has been discussed above. This fact is linked, on the one side, to the sequential way the platform has been released (via incremental updates) and it has been (as well it is expected to be further) exploited by partner organizations and end-users. On the other side, it is due to the web-based nature of the tool, which was designed having in mind the final goal of enabling an extremely user-friendly usage based on well know technologies and interfaces.

Thus, it can be immediately seen that the point of access to the platform is represented by a set of dynamic web pages generated using the PHP server-side scripting language[8] and delivered to the end-user's side by a web server (Apache HTTP Server[9], running on a MacOS X machine). Three kind of users are foreseen, i.e., project partner organizations (and their domain experts), job seekers (migrants) and employers (company recruiters, human resources managers, etc.), each with its own credentials and dedicated functionalities.

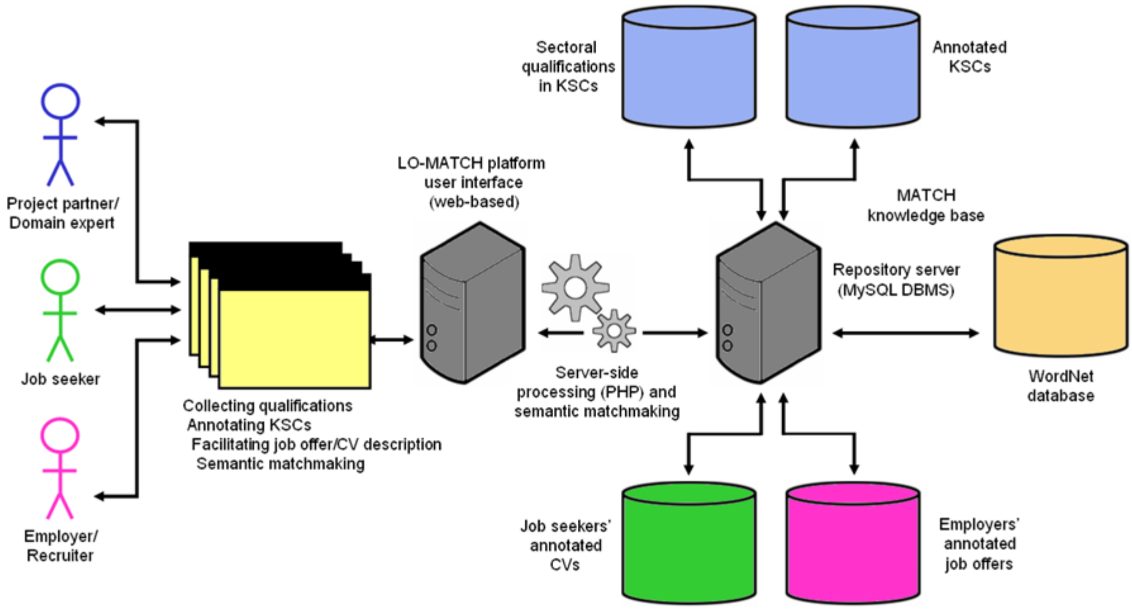

Figure 4.4. Software architecture of the LO-MATCH platform

As it will be shown in the following, specific interfaces are provided to let the various users intervene in the collection of educational and occupational profiles in EQF terms, in checking and possibly adjusting the computer-generated learning outcomes annotation, in inserting automatically annotated curriculum vitae and job offers and, lastly, depending on the perspective, in finding and ranking the best open positions a job seeker should apply for or the best candidates for a given job offer.

Data inserted and requested by job seekers and employers via their web browsers are managed by the PHP server-side logic via a dedicated repository server (developed as a MySQL database[10]), which is responsible for sharing the overall MATCH knowledge base. In particular, the server handles the collection of profiles (from the various sectors of interest) expressed in terms of knowledge, skills and competences, the WordNet database and annotated learning outcomes. Finally, the server-side

[8]http://php.net/

[9]http://httpd.apache.org/

[10]http://www.mysql.com/

logic is also responsible for implementing the search and sort algorithms required in the final matchmaking phase.

4.5 LO-MATCH functionalities and interfaces: semantic tools for job seekers and employers

As said, features are offered by the LO-MATCH platform on an account basis, that is, different tools are made available depending on the fact that the user currently logging in is a representative from a project partner organization, a job seeker or an employer (in the latter cases, the user should have been invited by a partner). In the following, the various tools will be discussed separately, by specifically focusing on their functionalities, on the underlying logic as well as on their interfaces.

4.5.1 Collecting occupational and educational profiles in EQF terms

The interface for profiles collection is targeted to domain experts from partner organizations. In fact, each organization was requested to login in the LO-MATCH platform and to insert educational and occupational profiles for the sectors that could be of interest based on the expertise of migrants they were planning to manage or the requirements of job offers they were going to support. Profiles are identified by the country of origin. They are structured in tasks and sub-tasks, each associated with a number of learning outcomes, expressed in terms of knowledge, skills and competences.

Information are inserted in English (which is considered the practical “vehicular language” actually enabling the link with the WordNet database), though translations in the partner’s national language could be added. Figure 4.5 shows different views of the LO-MATCH platform: the login page, the inserted profiles, the window for language translations and, finally, an excerpt of learning outcomes composing a profile.

It is worth observing that the platform was meant as a truly collaborative tool, since each partner was allowed to consult information inserted by other organizations, though it could only make modifications to its own data. While information is inserted, the platform splits sentences describing learning outcome in their constituting (key) words or groups of words, and performs an automatic annotation onto them based on the underlying ontological model. The annotation algorithm, which basically assigns WordNet synsets to relevant text items based on contextual cues, was designed to produce a redundant dataset (to be possibly adjusted in the next step), in order not to miss meanings possibly of interest for the user.

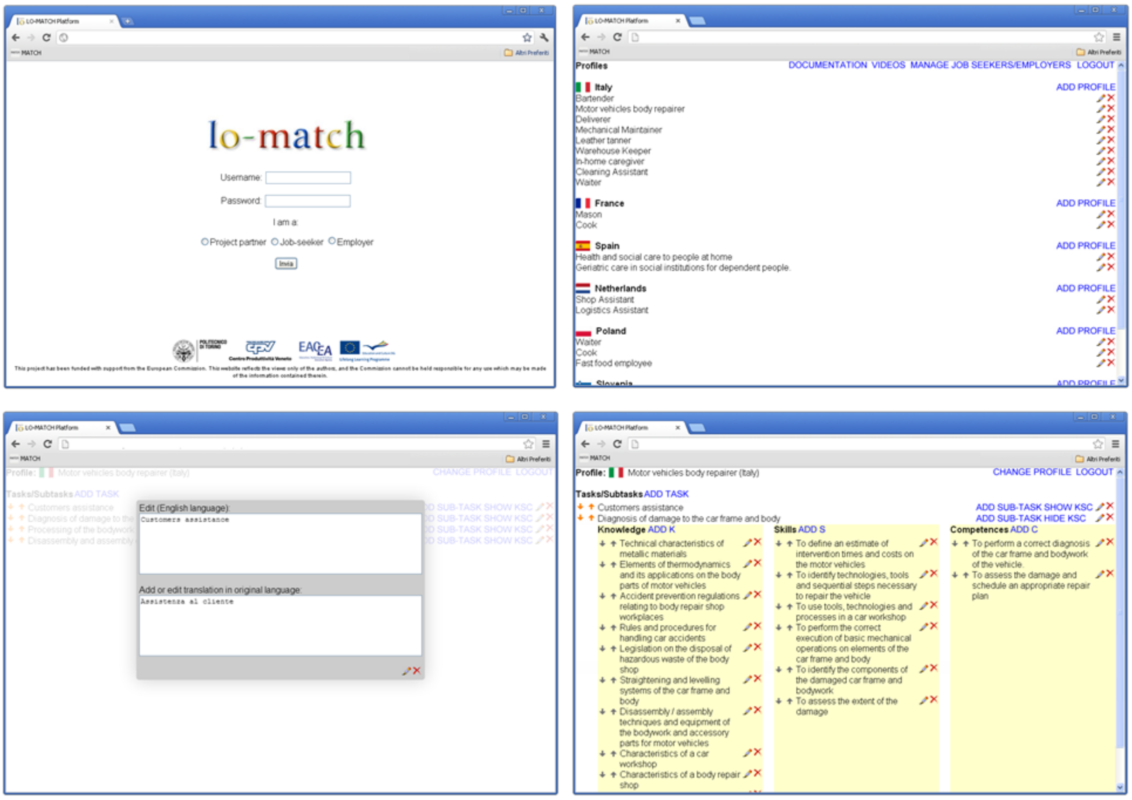

Figure 4.5. The LO-MATCH platform (project partner's view): login page (up--left), inserted profiles (up-right), language translations (bottom-left), learning outcomes composing a profile (bottom-right)

4.5.2 Checking the accuracy of automatic annotation

Thanks to the approach pursued, once the above data insertion step has been performed, end-users could immediately start using the platform for drafting their curriculum vitae and posting their job offers. Nonetheless, to guarantee improved accuracy in the overall matchmaking process and allow partner organizations to keep the control over the described computer-based pre-processing stage, a special tool for letting domain experts (without any particular knowledge of semantic technologies) perform a depth check on results achieved was developed. The basic idea was to release an easy-to-use graphics environment where results obtained by the machine could be possibly amended or integrated by trivial click and drag-and-drop operations. That is, as shown in the left part of Figure 4.6, the user can adjust action verbs, knowledge objects and context elements for annotated learning outcomes. Similarly, the user can operate on the WordNet-based annotation by adding/removing definitions assigned to words (and group of words) by choosing among those already available in the ontology or by inserting new terms and giving a definition for them.

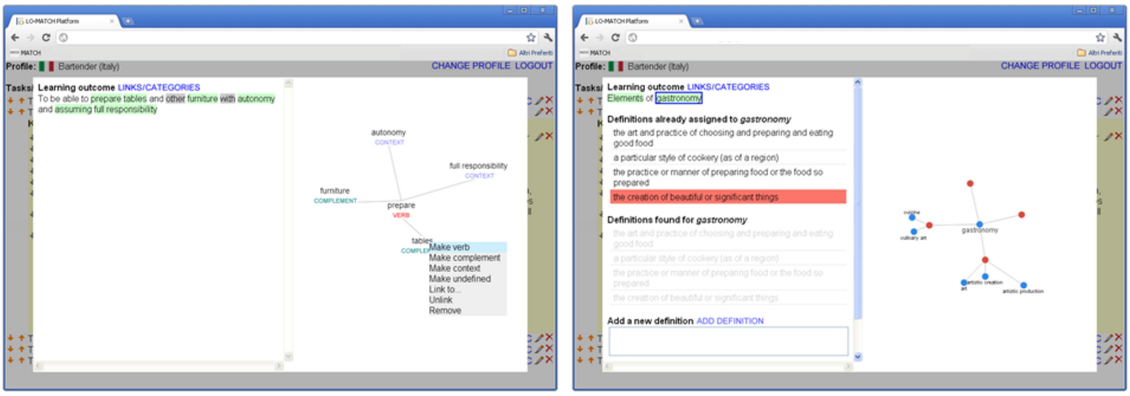

Figure 4.6. The LO-MATCH platform (project partner's view): adjusting KO, AV and CX (left) and choosing a definition for a concept (right)

In this case, as illustrated in the right part of Figure 4.6, textual information is accompanied by graphics representations, making the exploration of the underlying knowledge base more intuitive.

4.5.3 Inserting annotated curriculum vitae and job offers via seamless semantic facilitation

When a job seeker logs into the platform, he or she is provided with an interface designed for supporting the insertion of his or her learning acquirements. At the higher level, the interface is based upon the standard set in the Europass initiative for the Curriculum Vitae. That is, the user is first presented with a set of forms letting him or her indicate personal information, add a photo, specify basic details about work and well as education and training experiences, explicit linguistic abilities, etc.

Then, whereas in a common curriculum vitae, information e.g. about work experience is limited, for instance, to start and ending dates, occupation or position held, name of the employer, etc. and the main activities and responsibilities are generally described in a unstructured form (e.g., as free text), in the LO-MATCH platform a specific mechanism to let the job seeker elicit his or her requirements with an enhanced level of details by using the common language introduced above was defined.

That is, as shown in the bottom-left part of Figure 4.7, the user is allowed to link to each experience (i.e., to each formal, non-formal, or informal learning chance) a set of learning outcomes chosen by navigating the created knowledge base. But, more interestingly, he or she is allowed to carry out this "annotation" step in a simplified

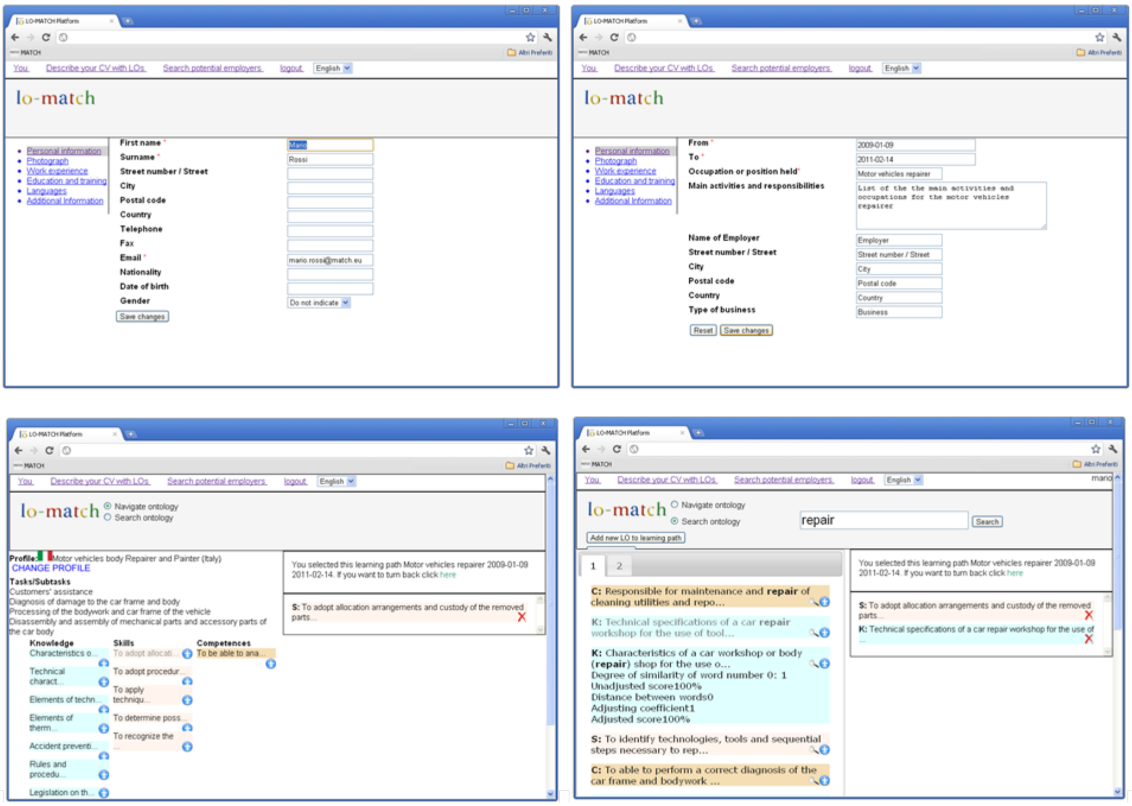

Figure 4.7. The LO-MATCH platform (job seeker's view): creation of a cv (up left and right), and selection of learning outcomes to be added to previous experience by navigating profiles (bottom-left) or by performing free text search (bottom-right)

way, by performing a free text search on the knowledge base and by getting hints about possible knowledge, skills or competences to include that are found based on semantic similarity (bottom right figure). The platform also suggests other elements to be possibly added, e.g., because of their relation to learning outcomes already selected, or because of their belonging to the same task, etc. The same strategy adopted for job seekers has been adapted and pursued also for letting employers publish annotated job offers in the system.

Compared to other platforms described in the research literature or available online, the advantage of the devised annotation approach is that, thanks to the automatic pre-processing stage, the end-user is not requested to have any knowledge of the underlying formal model (by the way, he or she is not even requested to have any knowledge of the concept of learning outcome). The devised interface let him or her search for elements that could best describe acquirements (in the case of a job seeker) or requirements (in the case of an employer) based on a simple query, like with any traditional search engine. Results obtained are then used to produce annotated curriculum vitae and/ or job offers that are based on a shared vocabulary relying upon standard concepts in the European lifelong learning domain like knowledge, skills and competences.

4.5.4 Computing the match between job offer and demand

In the above discussion it has been shown how, by pursuing the approach theorized in the research literature, semantic-based algorithms have been used in LO-MATCH to facilitate the creation of the annotated information involved in the overall e-job seeking and e-recruitment scenarios.

Afterwards, job seekers and employers can use again semantic-matching features embedded in the platform to obtain a ranked list of the best employment possibilities or candidates available for a given job offer, respectively (Figure 4.8). Nonetheless, it is worth observing that the user is allowed to setup the behaviour of the underlying similarity-based measures by using trivial configuration switches, thus passing from a pure keyword-based to a fully semantic-aware search (where the weights of the various relation types to overall score could be manually adjusted).

Following the guidelines that have been setup to support a wider usage and acceptance of semantic technologies, which ask for an enhancement of end-users' participation and awareness, motivations for the ranking produced by the platform (expressed as a similarity percentage) are provided by directly adding them in a well-known template. In fact, when the user is interested in getting more details about a given job position or candidate, he or she can explore information available by inspecting a template structured according to the Europass Curriculum Vitae and by consulting an intuitive graphics notation based on street light colours. For instance, a job seeker can check his or her acquirements against a particular job offer. For each learning outcome owned, a colour indicates whether it is also included among those required by the employer. When a match is found, either complete or incomplete, semantically similar elements are reported. The job seeker can also opt for a symmetrical representation, where required learning outcomes indicated in a given job offer are compared side-by-side with possessed ones.

Again, a comparable (though reverted) interface is available also for employers, thus allowing the LO-MATCH platform to address job matchmaking from the perspective of both parties involved.

4.6 Additional features: tag cloud-based representation of job offers and demands

During the development of the LO-MATCH platform, the exploitation of additional features allowing end-users to graphically compare curricula has been investigated.

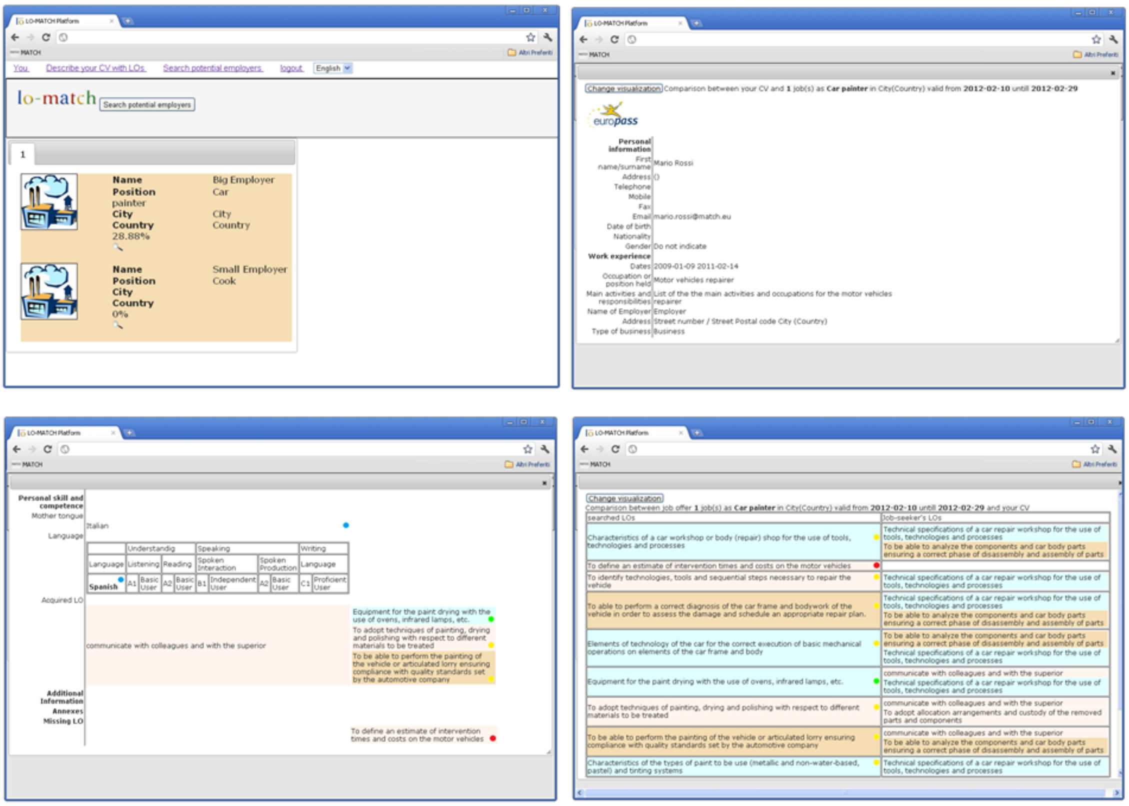

Figure 4.8. The LO-MATCH platform (job seeker's view): matching job offers (up-left), cv visualization (up-right), comparison of acquired and required learning outcomes (bottom left and right)

In particular, a prototype exploiting a tag cloud-based visualization technique to quickly depict information of interest in the selection and job seeking phases has been created; specifically, tag cloud characteristics, like font size and distance from the center of the cloud, are used to provide an overview of main characteristics of the abilities a given job seeker possesses or the suitability of a particular position, on the other hand, thus effectively supporting the comparison of qualifications and the job matchmaking both on the employer's and applicant's side.

It is worth remarking that this experimental feature has been proposed to end-user on a volunteer basis. The reason behind this choice rely on the fact that, in order to produce tag clouds, as it will further explained, end-users have to provide additional information, such as the degree of mastery/importance of a give knowledge, skill or competence. For this reason, it has been chosen to keep the matchmaking in the LO-MATCH platform as simple as possible, for standard end-users, and to propose tag cloud-based representations of demands and offers only to those volunteer end-users.

Tag clouds have been extensively used in different contexts. In fact, with the evolution of Web 2.0 and the opportunity for content providers and users to add metadata to published contents, a number of techniques have been developed to support users in performing search tasks, categorizing data and navigating the ever growing amount of information. In the above scenario, tag clouds started to be used as an attractive means for providing, at a first glance, a summary of the background information hidden into websites, blogs, and various online communities (like, for instance, Flickr, Delicious, etc.).

Basically, a tag cloud exploits effective information visualization techniques to present a visual overview of textual data, often corresponding to a set of tags. In a tag cloud, the font size used for drawing the tag is generally linked to importance (or frequency) of the tag itself. Originally, in tag clouds information was displayed using a rectangular line-by-line layout. Recently, the research community started studying the impact of other visual parameters on the attractiveness of tag cloud-based representations. As a matter of example, in [Luo et al. 2007] color information was included to visualize the actuality of tags. In [Rivadeneira et al. 2007], the impact of font weight and other text features on the execution of various user tasks was evaluated. A number of works dealt with the optimization of tag clouds layout. In [Shaw, 2005], the constraint of rectangular layouts was removed, and a graph-based structure was used to visualize relations between tags. In [Bielenberg and Zacher, 2005], a circular layout was proposed, and tag relevance was displayed by exploiting tag size as well as tag distance from the center of the cloud. In [Ad et al. 2010], tag placement based on similarity was exploited, by clustering similar tags in the cloud based on co-occurrence. A different approach was taken in [Vigas et al. 2009], where the basic tag cloud properties were considered with regard to aesthetic criteria.

Meanwhile, several studies were presented where the actual support provided by tag cloud-based representations to the execution of traditional tasks carried out on the web was analysed in both qualitative and quantitative terms. Though in some contexts (e.g., information mining) more trivial visualization techniques appeared to outperform tag clouds, in other scenarios encompassing visual browsing, multi-dimensional visualization, impression formation and information recognition/matching, tag cloud-based representations proved to be capable of providing a valuable support [Rivadeneira et al. 2007] [Hassan-Montero and Herrero-Solana, 2006].

In the experimental feature of the LO-MATCH platform, job applicant's characteristics as well as company's requirements have been represented by means of a tag cloud as follows: for each concept linked to a given learning outcome, a value of mastery/importance has been provided. The difference between degree of mastery and importance is linked to the particular kind of end user working on the platform: in fact, when a job applicant inserts a curriculum vitae in the knowledge base, he or she has to specify a degree of mastery, whereas when a company is inserting its requirements, it has to specify a degree of importance. In other words, the degree of mastery refers to the job offer perspective of the matchmaking process, whereas the degree of importance is related to the demand side. In particular, the degree of importance could assume the following values: low, medium-low, medium, medium-high, high.

Concepts stored in the knowledge base and their degree of importance/mastery are used to draw a cloud-based representation of a) the features of a job seeker's curriculum vitae better matching company's requirements, b) the main aspects of a company's working profile that could better valorize job position applicant's abilities. In the present implementation, the importance i of a concept is represented by means of the font size (with larger fonts indicating more relevant concepts), whereas the degree of mastery m is linked to the distance from the center of the cloud (e.g., for applicants with an exhaustive knowledge of the requested subjects, a compact tag cloud would be generated). This representation allows to simultaneously display both the dimensions of the matchmaking problem, i.e. company's needs and job seeker's characteristics. Thus, even non-skilled users/operators could easily see why a given matching has been obtained.

When focusing on the point of view of an employer searching a worker to hire, the font size used for drawing the tags is determined by sorting company's needs in a descending order based on importance i and by calculating the relative weight of a given concept with respect to the complete set of requirements. Then, concept coordinates are computed as:

$$x = r \times cos(\theta) \quad (4.3) \qquad\qquad y = r \times sin(\theta) \quad (4.4)$$

In such expressions, r is defined as

$$r = R \times (1 - m + D)/D \tag{4.5}$$

where
R is the maximum radius of the cloud
m is the degree of mastery
D is the number of possible values in the grading scale used for i and m
θ is a random angle.

The example reported in Table 4.2, presenting the requirements of a sample job position and the curricula of two possible applicants, should help to clarify the process. In particular, if values from 1 (low) to 5 (high) are used for measuring i and m (i.e., $D = 5$), concepts *product* and *selling techniques* would represent the 20%

of the knowledge requested by the company; then, *internal procedures and policies* and *health and safety rules* would represent the 12%; finally, the remaining concepts would be assigned the 4%. The font size would be determined by attributing a different value to the various percentage ranges, e.g., font size 10 for values between zero and 5%, etc. Then, assuming for instance $R = 500$ and choosing a random angle $\theta = 335°$, the *ICT* tag identified for the first applicant would be positioned at $x = 181.26$ and $y = -84.52$ (assuming the center of the cloud in $x = 0$ and $y = 0$).

Table 4.2. Degree of mastery for knowledge elements expressed by two job seekers and importance in the company's perspective

Knowledge element (concept)	**1st applicant**	**2nd applicant**	**Company**
Product	high	high	high
Selling techniques	-	-	high
Negotiation techniques	-	high	-
Customer identification techniques	-	high	-
Internal procedures and policies	low	medium-high	medium
Health and safety rules	medium	low	medium
ICT	medium-high	low	low
English	medium-high	low	low
Exposition techniques	low	medium-high	low
Organization techniques	-	-	low
Team working	low	medium-high	low
Basic sales legislation	low	low	low
Inventory techniques	-	-	low
Quality	low	medium-low	low
Analysis techniques	-	-	low

Figure 4.9 shows the tag clouds for the curricula of the two applicants, based on the taxonomy reported in Figure 4.10: since the company identified as a crucial aspect the knowledge of *product* and *selling techniques*, related tags are drawn with a large font, followed by the knowledge of *internal procedures and policies* and *health and safety rules*, and by several minor knowledge elements. The first applicant (Figure 4.9 left) has a high knowledge of the *product*, a medium-high knowledge of *English* and *ICT*, and a medium knowledge of *health and safety rules*. However, he or she has a low, or null, knowledge of other aspects of the work. Thus, only four elements are drawn close to the center of the cloud, whereas missing knowledge elements, like *selling techniques*, are placed on the external area (thus underlying their lack). In turn, the second applicant (Figure 4.9 right) already had some experience in the field; in fact, he or she shows a high knowledge of *product*, *negotiation*

techniques and *customer identification techniques*, a medium-high knowledge of several other aspects, and a low knowledge of remaining elements. Since, according to the ontology, *negotiation* and *customer identification techniques* are subsumed by the selling techniques concept, he or she possesses also a significant knowledge of *selling techniques*. Hence, the *product*, *selling techniques* and *internal procedures and policies* tags appear in the central area, thus making the second applicant the best (or, at least, a good) candidate for the given job.

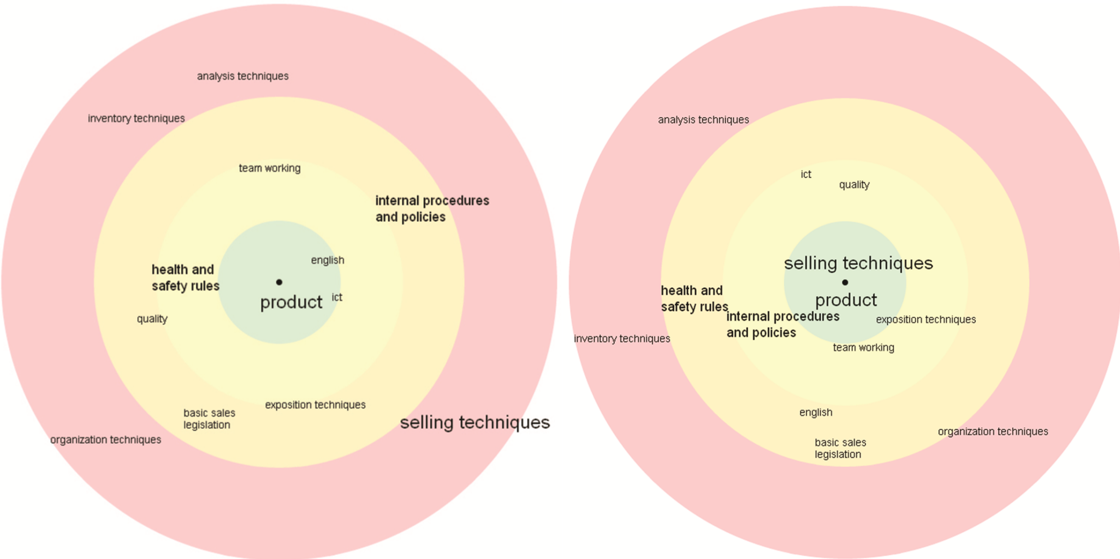

Figure 4.9. Tag cloud-based representation of the first applicant's curriculum vitae (left) and of the second applicant's curriculum vitae (right) in the company's perspective

The above examples analyse matchmaking results from the company's point of view. Nonetheless, comparable investigations could be carried out, for instance, from the perspective of job seekers, who are interested in finding companies that could recognize their abilities. The interface designed to this purpose is depicted in Figure 4.11 (still making reference to the example above). On the left hand side, a tag cloud shows how much the concepts expressed in the second applicant's résumé are made explicit in the description of the employers' requirements. In this case, in order to shift the focus on the applicant, the tag cloud is created by inverting i and m (i.e., by linking the font size and the distance from the center of the cloud to the degree of importance and the degree of mastery, respectively). On the right hand side, hints about those aspects the job seeker should address further in order to increase his or her opportunities of getting recruited by the given company are displayed: in this case, the candidate should improve his knowledge of health and safety rules (by raising it up to a medium level), and acquire some knowledge of

Figure 4.10. Portion of the taxonomy of interest for the tag clouds exemplified in Figure 4.9

organization techniques, inventory techniques and analysis techniques.

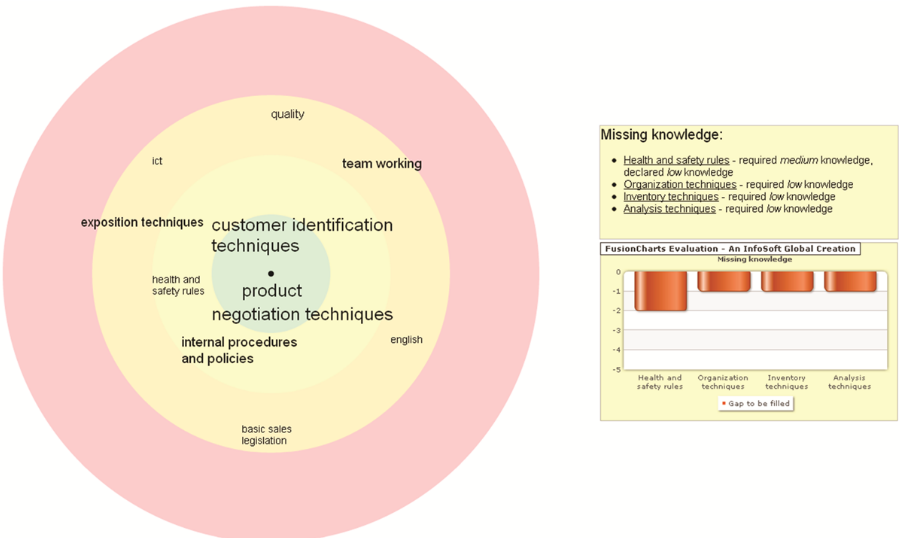

Figure 4.11. Tag cloud letting the second applicant (whose knowledge is reported in Table 4.2) compare his or her expertise with company's requirements

The job applicant could then exploit the devised platform to find a qualification (or part of it) providing the missing knowledge. In this case, the system would automatically record his requirements together with the needed level of importance, and would trigger the matchmaking with a demand input rather than with an offer description.

4.7 Conclusions

In this Chapter, the LO-MATCH platform, a platform developed during the current Ph.D. has been presented. This plaform represents a user-friendly solution facilitating job seekers and employers in the hard task of eliciting the acquirements got from their lifelong learning experiences as well as the requirements set to properly cover an open job position by using a shared common language and later supporting them in finding the best match between the above acquirements and requirements.

It is also the first work at modelling the underlying data structures that are necessary for dealing with the job matchmaking problem addressed by adhering to European tools and recommendations and by considering at the same time the expectations and needs of both actors involved.

By means of semantics, the comparison between offer and demand can be carried out by taking into account the inner details of each curriculum vitae and job posting and by processing the huge amount of information concerned by dealing with linguistic, cultural, systemic, etc. differences in descriptions as well as with ambiguity in words and sentences used by involved actors.

The approach used for managing the generally critical steps associated with the construction of the required knowledge base and for the annotation of job seeker and employer-provided information succeeded in reducing the effort to be put and knowledge required by project partner organizations and end-users.

In addition, a tag cloud-based graphical representation of the distance between acquirements and requirements has been presented.

The proposed platform has been exploited and tested by project partners, job seekers (migrants) and companies, with positive feedbacks.

Chapter 5

Conclusions

This thesis presents how semantics has been exploited to support the comparison of qualifications and job matchmaking. More specifically, the methodology followed within two European project - the TIPTOE and the MATCH projects - as well as the developed platforms (the TIPTOE and the LO-MATCH platforms) have been presented.

Both projects aimed at supporting students' and workers' mobility across Europe by means of:

- the development of a "common" European profile taking into account requirements of the labor world and outputs of the education and training domain, as in the case of the TIPTOE project. In this case, project partners exploited the platform for comparing 64 qualifications and job profiles, related to 4 professional figures of the trade sector. Each qualification, structured in tasks and subtasks, has been described in terms of knowledge, skills and competence elements. The TIPTOE platform allowed project partners to identify the common profile, by analysing concepts composing the description of learning outcomes, and by comparing them in order to find common elements even in those situations in which learning outcomes were described heterogeneously, both in terms of language, both at different levels of detail. At this purpose, four approaches have been identified and compared. Moreover, since partners previously specified, for each profile, its EQF level (a value between 1 and 8 providing information related to the fact that competences offered by a qualification or required for a given job profile are basic, operational, managerial, etc.) an additional feature allowed them to compute the EQF level for the whole common profile. Learning outcomes resulting from the comparison have then been collected and organised in the EQF ruler, a tabular representation of project results grouping knowledge, skills and competences that are required for each one of the 4 professional figures, both on the basis of task areas, both according to the EQF level;

- the development of a platform supporting job matchmaking, enabling the comparison of job offers and demands expressed according to EQF guidelines. Such a platform allows job seekers and recruiters to insert curricula and job offers structured in terms of learning outcomes. The platform performs an automatic annotation of inserted learning outcomes, that could be checked and validated by end-users (for example, to disambiguate terms with different meanings, etc.). In order to reduce end-users workload, it has been decided to pre-populate the platform with a set of most requested profiles, in the context of the project (migrants' employability), whose automatic annotation has been previously checked and validated by project partners.

  In this way, job seekers/recruiters could add to their curricula/job offers pre-annotated learning outcomes, both by browsing inserted profiles, both by performing a free text search. Moreover, they could also receive suggestions on learning outcomes to be potentially added in order to better describe their experience/request. Job offers and demands are then matched and ranked. Job seekers and recruiters willing to better understand the reasons behind a given result are provided with a comparative view, showing the gap between required and possessed learning outcomes. This way, they have an instrument suggesting them the more suitable corrective measures (for example, acquiring missing competences, in order to have more possibilities to be hired for a given job, or identifying candidates' weak points needing monitoring and reinforcement at the completion of the selection process). Finally, an additional feature exploiting a tag cloud-based representation of acquirements and requirements has been presented. By means of this depiction, exploiting the font size and the distance from the center of a target in order to communicate the importance of a requirement (or the degree of mastery), with respect to job seekers' (or job offers') characteristics, recruiters and job seekers could quickly compare required and acquired knowledge, skills and competences.

In the first case, in order to allow project partners to express job profiles and qualifications in a way as close as possible to natural language, their meaning was made explicit through the exploitation of an ad-hoc taxonomy, hierarchically structuring concepts contained in the description of learning outcomes, and grouping them in three trees, by distinguishing among knowledge objects, action verbs and context elements.

In this way, by browsing the taxonomy containing relations among concepts, similarities and differences among learning outcomes could be found even in those cases in which they were expressed by means of different terms.

Even though, on the one hand, this approach reduced partners' effort during the creation of the common profile, on the other hand, the construction of the taxonomy and the annotation of job profiles and qualifications became a time-consuming activity. For this reason, during the MATCH project, particular attention has been devoted to the creation of instruments enabling the automatic annotation. In fact, it has been chosen to use WordNet, a semantic thesaurus containing a considerable amount of concepts and relations among them. Inserted learning outcomes have then been broken in composing keywords (by also removing conjunctions, articles, etc. and by identifying the root of different terms), that have then been linked to WordNet concepts. The exploitation of WordNet laid the foundations for the use of this platform in different sectors, without limiting it to a specific one, as in the case of the TIPTOE project. In addition, the employment of a thesaurus potentially exploitable in different sectors made it possible that end-users (especially the recruiters) will no more need to be experts of a given sector in order to compare job offers and demands, for example, in order to know if a job seeker able to program in *Java* is a potential candidate for a job offer for an *object-oriented programmer* (difficult comparison if the human resources staff ignores that *Java* is an *object-oriented programming language*). Finally, it is worth remarking that the exploitation of EQF guidelines made this platform one of the first job matchmaking tools adopting European guidelines.

The experience acquired during the two projects above is currently being used within the TAMTAM "*Exploiting the TIPTOE plAtforM by transferring ECVET and EQF semAntic tools in a Multi-sectoral perspective*" project. In particular, in this project, a linguistic, geographical, systemic and sectoral transfer of the TIPTOE platform, by adopting the features for the automatic annotation exploiting WordNet developed within the MATCH project is being carried out.

Future research activities in the job matchmaking context will be devoted to link the LO-MATCH platform to existing platforms and repositories hosting profiles, curricula and job advertisements. Moreover, data coming from the existing qualification offer will be collected and matched against job seekers' missing competences, in order to provide end-users with a powerful tool supporting lifelong learning.

Bibliography

[Ad et al. 2010] I. Ad, K. Thiel, M.R. Berthold, "Distance Aware Tag Clouds", in Proc. of the IEEE Conf. on Systems Man and Cybernetics, pp. 2316–2322, 2010.

[Adomavicius and Tuzhilin, 2005] G. Adomavicius, A. Tuzhilin, "Toward the next generation of recommender systems: a survey of the state-of-the-art and possible extensions," IEEE Transactions on Knowledge and Data Engineering, pp. 734-749, 2005.

[Anderson et al. 2001] Anderson, L.W., Krathwohl, D.R., Bloom, B.S.: A taxonomy for learning, teaching, and assessing: a revision of Bloom's taxonomy of educational objectives. Long-man, (2001)

[Auer and Lehmann, 2007] Auer, S., Lehmann, J.: What have Innsbruck and Leipzig in common? Extracting Semantics from Wiki Content. Proceedings of European Semantic Web Conference (ESWC'07), LNCS 4519, 503–517, Springer (2007)

[Berners-Lee et al. 2001] T. Berners-Lee, J. Hendler, O. Lassila, "The semantic web," Scientific American, 248, pp. 35-43, 20011

[Bernstein et al. 2004] A. Bernstein, E. Kaufmann, A. Göhring, C. Kiefer, "Querying ontologies: a controlled English interface for end-users," in Proc. 4th International Semantic Web Conference, pp. 112-126, 2004

[Bielenberg and Zacher, 2005] K. Bielenberg, M. Zacher, "Groups in Social Software: Utilizing Tagging to Integrate Individual Contexts for Social Navigation", Master's Thesis, University of Bremen, 2005.

[Billig and Sandkuhl, 2002] A. Billig, K. Sandkuhl, "Matchmaking based on semantic nets: the XML-based approach of BaSeWeb," in Proc. 1st Workshop on XML Technologien fur das Semantic Web, pp. 39-51. 2002.

[Bizer et al. 2005] C. Bizer, R. Heese, M. Mochol, R. Oldakowski, R. Tolksdorf, E. Eckstein, "The impact of semantic web technologies on job recruitment processes," in Proc. Wirtschaftinformatik, 2005.

[Bloom, 1956] Bloom, B.S.: "Taxonomy of educational objectives - the classification of educational goals - Handbook I: cognitive domain"; Longman, (1956) [Bollegala et al. 2011] D. Bollegala, Y. Matsuo, M. Ishizuka, “A web search engine-based approach to measure semantic similarity between words,” IEEE Transactions on Knowledge and Data Engineering, 23:7, pp. 977-990, 2011.

[Bologna Declaration, 2001] “Bologna Declaration Communiqué of the meeting of European Ministers in charge of Higher Education (19 May 2001)”

[Bourque et al. 2003] Bourque, P., Buglione, L., Abran, A., April, A.: Bloom’s Taxonomy Levels for Three Software Engineer Profiles. Eleventh Annual International Workshop on Software Technology and Engineering Practice (STEP 2003), 123–129 (2003)

[Bourse et al. 2002] M. Bourse, M. Harzallah, M. Leclère, F. Trichet, “CommOnCV: modeling the competencies underlying a curriculum vitae,” in Proc. 14th International Conference on Software Engineering and Knowledge Engineering, pp. 65-73, 2002.

[Brown, 2004] D. Brown, “Unwanted online job seekers swamp HR staff,” Canadian HR Reporter, 17:7, pp. 1-2, 2004.

[Buckley et al. 2004] P. Buckley, K. Minette, D. Joy, J. Michaels, “The use of an automated employment recruiting and screening system for temporary professional employees: a case study,” Human Resource Management, 43:2-3, pp. 233-241, 2004

[Chen et al. 2008] Chen C.M., Peng C.J., Shiue J.Y.: “Ontology-based Concept Map for Planning Personalized Learning Path”; IEEE Conference on Cybernetics and Intelligent Systems (ICCIS 2008), 1337–1342 (2008)

[Cheetham and Chivers, 2005] G. Cheetham, G. E. Chivers, “Professions, competence and informal learning,” Edward Elgar Publishing, 2005.

[Colucci et al. 2003] S. Colucci, T. Di Noia, E. Di Sciascio, F. M. Donini, M. Mongiello, M. Mottola, “A formal approach to ontology-based semantic match of skills descriptions,” Journal of Universal Computer Science, 9:12, pp. 1437-1454, 2003.

[Colucci et al. 2006] S. Colucci, T. Di Noia, E. Di Sciascio, F. M. Donini, A. Ragone, R. Rizzi, “A semantic-based fully visual application for matchmaking and query refinement in B2C e-marketplaces,” in Proc. 8th International Conference on Electronic Commerce, pp. 734-749, 2006.

[Colucci et al. 2007] S. Colucci, T. Di Noia, E. Di Sciascio, F. M. Donini, A. Ragone, “Semantic-based skill management for automated task assignment and courseware composition,” Journal of Universal Computer Science, 13:9. pp. 1184-1211, 2007.

[Cubillos et al. 2006] Cubillos, C., Lamberti, F., Demartini, C.: “National Qualification Systems Integration Using Ontologies”, IFIP International Federation for Information Processing, vol 210, 285–289, 2006

[Cubillos et al. 2007a] Cubillos, C., Lamberti, F., Demartini, C.: “Integration at Vocational Education and Training Level through Mapped Ontologies”, Third

International Conference on Convergence and Hybrid Information Technology (ICCIT2008), vol 1, 117–122 (2008)

[Cubillos et al. 2007b] Cubillos, C., Lamberti, F., Demartini, C.: "Integrating European Qualification Systems with OWL Ontologies", Electronics, Robotics and Automotive Mechanics Conference (CERMA 2007), 752–757, 2007

[Cubillos et al. 2008] C. Cubillos, F. Lamberti, C. Demartini, "Integration at vocational education and training level through mapped ontologies," in Proc. 3rd International Conference on Convergence and Hybrid Information Technology, pp.117-122, 2008.

[De Coi et al. 2007] J. L. De Coi, E. Herder, A. Koesling, C. Lofi , D. Olmedilla , O. Papapetrou , W. Siberski, "A model for competence gap analysis," in Proc. Conference on Web Information Systems and Technologies, 2007

[Dexter and Davies, 2009] Dexter, H., Davies, I.: "An ontology-based curriculum knowledgebase for managing complexity and change"; 9th IEEE International Conference on Advanced Learning Technologies, 136–140 (2009)

[Earl, 2001] M. J. Earl, "Knowledge management strategies: toward a taxonomy," Journal of Management Information Systems, 18:1, pp. 215-242, 2001.

[Fazel-Zarandi and Fox, 2009] M. Fazel-Zarandi, M.S. Fox, "Semantic matchmaking for job recruitment: an ontology-based hybrid approach," in Proc. 3rd International Workshop on Service Matchmaking and Resource Retrieval, 2009.

[Fazel-Zarandi and Fox, 2010] M. Fazel-Zarandi, M. S. Fox, "Reasoning about skills and competencies," in Proc. IFIP WG 5.5 Working Conference on Virtual Enterprises, pp. 372-379, 2010.

[Fazel-Zarandi et al. 2011a] M. Fazel-Zarandi, H. J. Devlin, Y. Huang, N. Contractor, "Expert recommendation based on social drivers, social network analysis, and semantic data representation," Proc. 2nd International Workshop on Information Heterogeneity and Fusion in Recommender Systems, pp. 41-48, 2011.

[Fazel-Zarandi and Fox, 2011b] M. Fazel-Zarandi, M. S. Fox, "Constructing expert profiles over time for skills management and expert finding," in Proc. 11th International Conference on Knowledge Management and Knowledge Technologies, 2011.

[Fazel-Zarandi and Fox, 2012] M. Fazel-Zarandi, M.S. Fox, "An ontology for skill and competency management," in Proc. 7th International Conference on Formal Ontology in Information Systems, 2012.

[Fazel-Zarandi et al. 2013] M. Fazel-Zarandi, M.S. Fox, E. Yu, "Ontologies in expertise finding systems: modelling, analysis, and design," Ontology-Based Applications for Enterprise Systems and Knowledge Management, pp. 158-177, 2013.

[Fernandez et al. 1997] M. Fernandez, P.A. Gomez, N. Juristo, "Methontology: from ontological art toward ontological engineering," in Proc. Spring Symposium Series on Ontological Engineering, pp. 33-40, 1997.

[Ferreira and Filipe, 2009] Ferreira, J.C. and Filipe, P.: "Academic Ontology to Support the Bologna Mobility Process"; 2nd International Conference on Adaptive Science & Technology (ICAST 2009), 308–313 (2009)

[Gatteschi et al. 2009] Gatteschi, V., Lamberti, F., Salassa, F., Demartini, C.: "An Automatic Tool Supporting Life-long Learning Based on a Semantic-Oriented Approach for Comparing Qualifications"; IADIS International Conference on Cognition and Exploratory Learning in Digital Age (CELDA 2009), 354–358 (2009)

[Gatteschi et al. 2010] V. Gatteschi, F. Lamberti, A. Sanna, C. Demartini, "A semantic matchmaking system for job recruitment," in Proc. I-KNOW Conference, 2010.

[Gatteschi et al. 2011] V. Gatteschi, F. Lamberti, A. Sanna, C. Demartini, "A ranking tool exploiting semantic descriptions for the comparison of EQF-based qualifications," Journal of Universal Computer Science, vol. 17:7, pp. 1060-1077, 2011.

[Gestwicki, 2008] Gestwicki, P.: Work in Progress - Curriculum Visualization. Frontiers In Education Conference (FIE 2008), T3E13-T3E14 (2008)

[Gómez-Pérez et al. 2007] A. Gómez-Pérez, J. Ramírez, B. Villazón-Terrazas, "Reusing human resources management standards for employment services," in Proc. 1st Industrial Results of Semantic Technologies, 2007.

[Gruber, 1993a] T. R. Gruber, "A translation approach to portable ontology specifications," Knowledge Acquisition, 5:2, pp. 199-220, 1993.

[Gruber, 1993b] Gruber, T.R.: "Toward principles for the design of ontologies used for knowledge sharing"; Padua workshop on Formal Ontology, 1–23 (1993)

[Guarino, 1998] N. Guarino, "Formal ontology and information systems," Proc. International Conference on Formal Ontology in Information Systems, pp. 3-15, 1998.

[Hackelbusch, 2006] Hackelbusch, R.: "Handling Heterogeneous Academic Curricula: National qualification systems integration using ontologies"; 17th Int. Workshop on Database and Expert Systems Applications (DEXA 2006), 344–348 (2006)

[Hansen et al. 2010] D. L. Hansen, T. Khopkar, J. Zhang, "Recommender systems and expert locators," Encyclopaedia of Library and Information Sciences, Third Edition, pp. 4433-4441, 2010.

[Hassan-Montero and Herrero-Solana, 2006] Y. Hassan-Montero, V. Herrero-Solana, "Improving Tag-Clouds as Visual Information Retrieval Interfaces", in Proc. of the Int. Conf. on Multidisciplinary Information Sciences and Technologies, 2006.

[Hoffman, 2008] Hoffmann, M.: "Using Bloom's Taxonomy of learning to make engineering courses comparable"; EAEEIE 19th Annual Conf. (2008), 205–209 (2008)

[IEEE and ACM, 2001] IEEE Computer Society and Association for Computing Machinery "Computing Curricula 2001 - Computer Science", 2001

[Ionescu et al. 2012] B. Ionescu, I. Ionescu, V. Florescu, A. Tinca, “Semantic annotation and association of web documents: a proposal for semantic modelling in the context of e-recruitment in the IT field,” Accounting and Management Information Systems, 11:1, pp. 76-96, 2012.

[Kessler et al. 2008] R. Kessler, N. Bechet, E. Roche, M. El-Beze, J.M. Torres-Moreno, “Automatic profiling system for ranking candidates answers in human resources”, in Proc. On the Move to Meaningful Internet Systems Workshop, pp. 625-634, 2008.

[Koper, 2004] R. Koper, “Use of the semantic web to solve some basic problems in education: increase flexible, distributed lifelong learning; decrease teacher’s workload,” Journal of Interactive Media in Education, 6, 2004.

[Lee, 2011] I. Lee, “Modelling the benefit of e-recruiting process integration,” Decision Support Systems, 51:1, pp, 230-239, 2011.

[Legge, 2005] K. Legge, “Human resource management: rhetorics and realities,” Palgrave Macmillan, 2005.

[Luo et al. 2007] B.Y.L. Luo, T. Hentrich, B.M. Good, M.D. Wilkinson, “Tag Clouds for Summarizing Web Results”, in Proc. of the 16th Int. Conf. on World Wide Web, pp. 1203–1204, 2007.

[Lv and Zhu, 2006] H. Lv, B. Zhu, “Skill ontology-based semantic model and its matching algorithm”, in Proc. 7th International Conference on Computer-Aided Industrial Design and Conceptual Design, pp. 1-4, 2006.

[Lv and Zhu, 2010] H. Lv, B. Zhu, “Elastic information matching technology and its application in electronic recruitment,” in Proc. 11th International Conference on Computer-Aided Industrial Design and Conceptual Design, pp. 1582-1585, 2010.

[Mirizzi et al. 2009] Mirizzi, R., Di Noia, T., Di Sciascio, T.: “A Semantic Web enabled System for Résumé Composition and Publication”; IEEE International Conference on Semantic Computing (ICSC 2009), 583–588, 2009

[Mirizzi et al. 2010] R. Mirizzi, A. Ragone, T. Di Noia, E. Di Sciascio, “Semantic tag cloud generation via DBpedia,” in Proc. E-Commerce and Web Technologies, pp. 36-48, 2010.

[Mochol et al. 2004] M. Mochol, R. Oldakowski, R. Heese, “Ontology based recruitment process,” in Proc. Informatik Verbindet, 2004.

[Mochol et al. 2007a] M. Mochol, H. Wache, L. Nixon, “Improving the accuracy of job search with semantic techniques,” in Proc. 10th International Conference on Business Information Systems, pp. 301-313, 2007.

[Mochol et al. 2007b] M. Mochol, A. Jentzsch, J. Euzenat, “Towards a methodology for selection suitable matching approaches: a case study,” in Proc. 4th European Semantic Web Conference, 2007.

[Mochol et al. 2007c] M. Mochol, A. Jentzsch, H. Wache, "Suitable employees wanted? Find them with semantic techniques," in Proc. Workshop on Making Semantics Web For Business at European Semantic Technology Conference, 2007.

[Monster, 2012] Monster. Monster Deutschland and TMP Worldwide: Recruiting Trends 2012.

[NIESR, 2011] National Institute of Economic and Social Research (NIESR), "Labour mobility within the EU -The impact of enlargement and the functioning of the transitional arrangements", 2011. http://www.niesr.ac.uk/pdf/050811_152043.pdf Cited February 2013

[Nemirovskij et al. 2007] Nemirovskij, G., Egner, M.T., Heuel, E.: "SWAPS: Semantic Web Approach for Personalisation of Study"; 7th IEEE International Conference on Advanced Learning Technologies (ICALT 2007), 711–712 (2007)

[Netcraft, 2011] Netcraft Web Server Survey. http://news.netcraft.com/archives/category/web-server-survey/. Cited February 2013

[Ohler, 2008] Ohler, J.: The SemanticWeb in Education - What happens when the read-write web gets smart enough to help us organize and evaluate the information it provides? EDUCAUSE Quarterly Magazine, 31 (4), 7–9 (2008)

[Pernici et al. 2006] Pernici, B., Locatelli, P., Marinoni, C.: "The eCCO System: An eCompetence Management Tool Based on Semantic Networks"; Workshop on Ontology Content and Evaluation in Enterprise (OnToContent 2006), 1088–1099, 2006

[Ployhart, 2006] R. E. Ployhart, "Staffing in the 21st century: New challenges and strategic opportunities," Journal of Management, 32:6, pp. 868-897, 2006.

[Poyry et al. 2002] Poyry, P., Pelto-Aho, K., Puustjarvi, J.: "The role of metadata in the CUBER system"; Proc. of the 2002 annual research Conf. of the South African institute of computer scientists and information technologists on Enablement through technology, 16–18 (2002)

[Poyry and Puustjarvi, 2003] Poyry, P., Puustjarvi, J.: "CUBER: a personalised curriculum builder"; Proceedings of the 3rd IEEE Int. Conf. on Advanced Learning Technologies, 326–327 (2003)

[Radevski and Trichet, 2006] V. Radevski, F. Trichet, "Ontology-based systems dedicated to human resources management: an application in e-recruitment," in Proc. On the Move to Meaningful Internet Systems Workshop, pp. 1068-1077, 2006.

[Rafter and Smyth, 2001] R. Rafter, B. Smyth, "Passive profiling from server logs in an online recruitment environment," in Proc. IJCAI Workshop on Intelligent Techniques for Web Personalisation, 2001.

[Reich et al. 2002] J. R. Reich, P. Brockhausen, T. Lau, U. Reimer, "Ontology-based skills management: goals, opportunities and challenges," Journal of Universal Computer Science, 8:5, pp. 506-515, 2002.

[Rivadeneira et al. 2007] A.W. Rivadeneira, D.M. Gruen, M.J. Muller, D.R. Millen, "Getting Our Head in the Clouds: Toward Evaluation Studies of Tag Clouds", in Proc. of the SIGCHI Conf. on Human Factors in Computing Systems, pp. 995–998, 2007.

[Ronchetti and Sant, 2007] Ronchetti, M., Sant, J.: "Curriculum Management and Review: an ontology-based solution"; World Conference on E-Learning in Corporate, Government, Healthcare, and Higher Education (ELEARN 2007), 6476–6482 (2007)

[Shaw, 2005] B. Shaw, "Utilizing Folksonomy: Similarity Metadata from the del.icio.us System", Project Proposal, December 2005.

[Saini and Ronchetti, 2003] Saini P., Ronchetti M.: Deriving ontology-based metadata for e-learning from the ACM Computing Curricula. IADIS International Conference e-Society 2003, 892–898 (2003)

[Sampson and Fytros, 2008] Sampson, D., Fytros, D.: Competence Based Educational Metadata for Supporting Lifelong Competence Development Programmes. The 8th IEEE International Conference on Advanced Learning Technologies (ICALT 2008), 288–292 (2008)

[Sommaruga and Catenazzi, 2007] Sommaruga, L., Catenazzi, N.: Curriculum visualization in 3D. 12th International Conference on 3D Web Technology, 177–180 (2007)

[Spivey, 2007] Spivey, G.: A Taxonomy for learning, teaching, and assessing Digital Logic Design. Frontiers In Education Conference - Global Engineering: Knowledge Without Borders, Opportunities Without Passports (FIE 2007), F4G-9-F4G-14 (2007)

[Staab and Studer, 2004] Staab, S. and Studer, R.: Handbook on Ontologies. International handbooks of Information Systems, Springer (2004)

[Starr et al. 2003] Starr, C.W., Manaris, B.Z., Stalvey, R.H.: Bloom's taxonomy revisited: specifying assessable learning objectives in computer science. SIGCSE (2008), 261–265 (2008)

[Tang, 2009] Tang, A.: "An Ontological Approach to Curriculum Development"; International Conference on Engineering Education (ICEED 2009), 219–224 (2009)

[Tao et al. 2005] Tao, F. B., Millard, D., Davis, H. and Woukeu, A.: Managing the Semantic Aspects of Learning using the Knowledge Life Cycle. The 5th IEEE International Conference on Advanced Learning Technologies (ICALT 2005), 575–579 (2005)

[Taubner and Brössler, 2000] D. Taubner, P. Brössler, "The people make the project," In Proc. ESCOP-SCOPE Conference, pp. 105-113, 2000.

[Tiropanis et al. 2009] Tiropanis, T., Davis, H., Millard, D. and Weal, M.: Semantic Technolologies for Learning and Teaching in the Web 2.0 Era. Intelligent Systems, IEEE, 24 (6), 49–53 (2009)

[Trichet, 2004] Trichet, F., "Human resource management and semantic web technologies," in Proc. International Conference on Information and Communication Technologies: From Theory to Applications, pp. 641-642, 2004.

[Vega, 1990] J. Vega, "Semantic matching between job offers and job search requests," Proc. 13th Conference on Computational Linguistics, pp. 67-69, 1990.

[Vigas et al. 2009] F.B. Vigas, M. Wattenberg, J. Feinberg, "Partecipatory Visualization with Wordle" in IEEE Transactions on Visualization and Computer Graphics, vol. 15, no. 6, 2009.

[Yahiaoui et al. 2006] L. Yahiaoui, Z. Boufaïda, Y. Prié, "Semantic annotation of documents applied to e-recruitment," in Proc. Semantic Web Applications and Perspectives Conference, 2006.

[Yimam-seid and Kobsa, 2003] D. Yimam-seid, A. Kobsa, "Expert finding systems for organizations: problem and domain analysis and the DEMOIR approach," Journal of Organizational Computing and Electronic Commerce, 13:1, pp. 1-24, 2003.

[Younker, 1998] E. Younker, "Skills management: the linchpin for IT work force management," InSide GartnerGroup, 1998.

[Zhong et al. 2002] J. Zhong, H. Zhu, J. Li, Y. Yu, "Conceptual graph matching for semantic Search," in Proc. of the International Conference on Computational Science, 2002.