

LO-MATCH: A semantic platform for matching migrants' competences with labour market's needs

*Original*

LO-MATCH: A semantic platform for matching migrants' competences with labour market's needs / Gatteschi, Valentina; Lamberti, Fabrizio; Demartini, Claudio Giovanni. - STAMPA. - 1:(2012), pp. 994-998. (Intervento presentato al convegno 3rd IEEE Global Engineering Education Conference (EDUCON2012) tenutosi a Marrakesh (Morocco) nel 17-20 April, 2012) [10.1109/EDUCON.2012.6201168].

*Availability:*

This version is available at: 11583/2466779 since:

*Publisher:*

*Published*

DOI:10.1109/EDUCON.2012.6201168

*Terms of use:*

This article is made available under terms and conditions as specified in the corresponding bibliographic description in the repository

*Publisher copyright*

(Article begins on next page)

This is an electronic version (author's or accepted version) of the paper:

Gatteschi V., Lamberti F., Demartini C.,  
“LO-MATCH: A semantic platform for matching migrants' competences with labour market's needs,”  
IEEE Global Engineering Education Conference (EDUCON), pp. 1-5, 2012.

DOI: 10.1109/EDUCON.2012.6201168

Link to IEEE Xplore®: <http://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=6201168>

© 2012 IEEE. Personal use of this material is permitted. Permission from IEEE must be obtained for all other uses, in any current or future media, including reprinting/republishing this material for advertising or promotional purposes, creating new collective works, for resale or redistribution to servers or lists, or reuse of any copyrighted component of this work in other works.

# LO-MATCH: A Semantic Platform for Matching Migrants' Competences with Labour Market's Needs

Valentina Gatteschi, Fabrizio Lamberti and Claudio Demartini

Dipartimento di Automatica e Informatica

Politecnico di Torino

Torino, Italy

valentina.gatteschi@polito.it, fabrizio.lamberti@polito.it, claudio.demartini@polito.it

**Abstract**—Citizens' mobility and employability are receiving ever more attention by the European legislation. Various instruments have been defined to overcome lexical and semantic differences in the descriptions of qualifications, résumés and job profiles. However, the above differences still represent a significant constraint when abilities of non-European people have to be validated either for education and training or occupation purposes. In this work, a web platform that exploits semantic technologies to address such heterogeneity issues is presented. The platform allows migrants to annotate their knowledge, skills and competences in a shared format based on the European tools. The resulting knowledge base is then used to enable the automatic matchmaking of job seekers' abilities with companies' needs. The platform can additionally be used to support students and workers in the identification of their competence gap with respect to a given education or occupation opportunity, so that to personalize their further training.

**Keywords-component;** *ontologies, competences, lifelong learning, WordNet*

## I. 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 (EQF) [2], the European Credit Transfer System (ECTS) [3], 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 their abilities could be best valorized), 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, such as ontologies (explicit specifications of a conceptualization, as defined in [4]) and taxonomies (classifications arranged in a hierarchical structure), 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, in this work a strategy to tackle heterogeneity issues in the descriptions of qualifications, résumés and labor market's needs due to the use of non-shared vocabularies is proposed. In particular, the development of the LO-MATCH platform, a semantic tool designed to support the identification/development of required/missing competences currently under development in the framework of the MATCH "Informal and non-formal competences matching device for migrants' employability and active citizenship" project, is presented. This tool exploits the general purpose WordNet [5] ontology 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.). Moreover, with LO-MATCH, owned and expected skills can be annotated according to a shared format based on European recommendations.

The rest of paper is organized as follows: Section 2 gives an overview of recent works exploiting semantics for the classification and comparison of educational and occupational profiles, while Section 3 presents the structure of the WordNet semantic thesaurus. The objectives of the MATCH project as well as the steps that are being carried out for developing the LO-MATCH platform are devised in Section 4. Finally, conclusions are presented in Section 5.

## II. BACKGROUND

Several research activities had been already carried out in the direction of classifying and comparing learning outcome elements in education and occupation profiles. [6] – [7] started from the assumption that the information about training courses that is available on the internet is usually unstructured, thus making comparison of study courses difficult to perform, and proposed an approach based on metadata (data about data) to describe courses provided by European Universities in order to support learners looking for higher education courses that match their needs. However, a software system such as the one just presented could be able to understand the structure of a training course, but not its content. A solution to overcome this limitation is presented in [8]: in this work the authors proposed the exploitation of the Bloom's taxonomy [9], a taxonomy identifying six levels of learning mastery (knowledge, comprehension, application, analysis, synthesis, and evaluation), in order to compare competences acquired by new graduates, graduates with four years of experience, and experienced software engineers by assigning them to one of the above categories. Another interesting research is reported in [10]: here the author defined a quantitative approach, based on the Bloom's taxonomy, to compute the level of skills belonging to a qualification and shown how three-dimensional graphs putting into relations the knowledge dimension and the cognitive dimension together with a degree of mastery could support the comparison of Engineering courses. While the Bloom's taxonomy is mostly used to classify skills, other works suggest new strategies to compare different elements of a qualification (e.g. the knowledge taught in a course, the context in which the skill is applied, etc.). [11] proposed a methodology for comparing occupational and educational profiles belonging to the trade sector in order to identify their common elements and, thus, to create a common profile taking into account the requirements of both fields. The methodology proposed by the authors consisted in linking the elements belonging to occupational and educational descriptions to a set of concepts, organized into an ad-hoc taxonomy, in order to exploit relations among terms and thus, to allow a software platform to compare learning outcomes denoting the same meaning even in those cases in which they are expressed by means of different sentences. [12] performed a similar work in job seeking and recruitment fields: in particular the authors presented an intelligent recruitment platform based on an ad-

hoc ontology to help companies to find the most suitable applicant for a given job and vice versa.

While the above works focus on strategies to compare training courses and curricula on the basis of a keyword-based search or by exploiting an ontology that could be applied to few sectors [11][12] or to few aspects of a qualification [8][10], this work presents how a general purpose ontology, such as the WordNet lexical database could be used to compare qualifications, résumés and job offer, thus supporting mobility of students and workers in different sectors and contexts.

## III. WORDNET

WordNet is a large lexical database of English collecting words (i.e. nouns, verbs, adjectives and adverbs) and grouping them in a set of synonyms called *synsets*. For each synset, a short general definition of the group of terms to which it is linked is provided, together with semantic relations with other synsets. The advantage of such a database consists in the fact that it provides not only a combination of dictionary and thesaurus that is more intuitively usable, but it also supports automatic text analysis and artificial intelligence applications. Moreover, apart from grouping words together based on their meanings, it has some additional features that makes it more than just a thesaurus: first, it connects not only word forms (i.e. strings of letters) but also the specific meaning of terms; second, it provides information about the type of semantic relation among words. In a thesaurus, instead, groups of terms are only created on the basis of meaning similarity. The WordNet semantic thesaurus could be browsed online (by using the official application or one of the free software tools developed in related projects), or could be downloaded and exploited for personal use or software applications.

The structure of the WordNet semantic thesaurus is the following: synonyms (that are the more frequent relations among words) are grouped into synsets. At the present time, around 117 000 synsets have been collected and interlinked. Each synset is then linked to other synsets through a given relation and for each synset a brief description, together with one or more sentences showing the use of the words belonging to the synset, is provided. Most common 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. A less frequent, but still important relation is meronymy *X is a meronym of Y if X is a part/member of Y*. Other relations are troponymy (the presence of a "manner" relation between two synsets, like in communicate-talk-whisper, where the manner depends on the volume) for verbs, and antonymy (like wet-dry) for adjectives, etc. However, since a full analysis of WordNet semantic relations is out of the scope of this work, the reader could find more information on them on the WordNet website.

## IV. THE LO-MATCH PLATFORM

The objective of the MATCH project is to support the recognition of skills, competences and qualifications of migrant workers in Europe by the creation of a web-based functional tool (the LO-MATCH software platform) connecting the migrants' competences acquired in formal, non-formal and

informal contexts to occupational profiles and to companies' labour demand. Such a tool should be able to:

- allow migrants 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' 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 learners and workers in the identification of training courses they could attend in order to obtain missing competences and, thus, to increase their possibility to be employed by a given company. In this view, migrants' résumés are matched with existing training offers.

According to the above requirements, the LO-MATCH platform should be able to compare résumés, job offers and qualifications. However, since these three elements could be expressed by means of a huge heterogeneity of terms, the exploitation of tools developed within the Semantic Web initiative could increase the efficacy of comparisons and improve matchmaking. In order to better clarify the above statement, let us consider a migrant able to *develop Java applications* and a company looking for someone having the ability to *program with object-oriented languages*. While the similarity of demand and offer is clear for a human being (since Java is an object-oriented programming language), for a computer software it could be difficult to understand that the above sentences denote the same thing. On the other hand, if the above skills are annotated according to a shared ontology, by specifying relevant concepts (e.g. for the candidate's skill terms such *Java* and *(to) develop* could be selected, while for the company's request words like *(to) program* and *object-oriented language* could be specified), a machine could be able to browse the ontology and to understand that the concepts depicting offer and demand are somehow related (e.g. by an "is-a" relationship for *Java* and *object-oriented language*, or by a synonym relation, for what it concerns the *develop* and *program* concepts). Hence, at the beginning of the MATCH project, existing ontologies have been investigated and the WordNet semantic thesaurus has been selected for annotating résumés, job offers and qualifications, since it turned out to

have a large set of concept and relations among them as well as a considerable community of users and applications. Moreover, the *synset – definition* structure proves to be extremely useful in those cases in which a word could assume different meanings (e.g. *Java* could be a programming language, a type of coffee, an island, etc). In this view, in the LO-MATCH platform, each time a job seeker (or a company, or training institute) wants to insert his offer (or demand), he/she should also specify the concepts by means of which he/she would annotate the offered (or needed, or provided) characteristics, and he/she should chose, among their definition, the more suitable one.

However, since the actors that will use the LO-MATCH platform should probably familiarize with an ontology like WordNet, before being able to fully exploit its potential, it has been decided to initially collect and annotate in the LO-MATCH knowledge base a set of (most requested and offered) professional figures/qualifications and to develop a tool for automatic annotation of résumés, job offers and qualifications by referring them to the already inserted information. In this way, each time a company (or job seeker, or training institute) writes a sentence to characterize its request (or offer), the system will identify terms composing the sentence, and will browse the WordNet ontology in order to find an already annotated element of a professional figure/qualification. This way the platform could be immediately used like any other recruitment portal, thus loyalizing end-users. In a second phase, however, job seekers, companies and training institutes will be able to voluntarily perform the annotation themselves.

The phases for the development of the first release of the LO-MATCH platform are: a) collection of professional figures/qualifications; b) their annotation in the knowledge base; c) testing with a set of résumés, job advertisements and qualifications; d) development of additional features and dissemination.

#### A. Collection of professional figures/qualifications

In this phase, project partners (that are Chamber of Commerce, Training Authorities, Vocational Education and Training providers associations, etc.) investigate their country's employment situation in order to identify professional figures and qualifications to be inserted in the LO-MATCH knowledge base. This step is currently in progress, and it is been carried out by having in mind the project's objective, thus, by trying to

The screenshot shows a user interface for a professional profile. At the top, it says 'Profile: Bartender (Italy)' and 'CHANGE PROFILE LOGOUT'. Below this, there are four main sections: 'Tasks/Subtasks', 'Knowledge', 'Skills', and 'Competences'. Each section has a list of items with small icons (up/down arrows and a checkmark) next to them. At the bottom right, there are three buttons: 'ADD SUB-TASK SHOW KSC', 'ADD SUB-TASK HIDE KSC', and 'ADD SUB-TASK SHOW KSC'.

Figure 1. Learning outcomes related to the *to prepare bar services* task, belonging to an Italian Bartender professional figure

find which are the most requested professional figures, or which are the sectors that could benefit at most from the exploitation of the LO-MATCH platform. The result of this phase is 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 is collected by means of desk research or interviews with relevant stakeholders.

Collected information is 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 [13]. In this work the authors suggested a way to cope with EQF guidelines by introducing three families of elements: knowledge objects (KO), action verb (AV) and context (CX) concepts. According to [13], a knowledge could be seen as a set of KO, a skill could be defined as a KO "put into action" through an AV, hence by one or more pairs KO - AV, and a competence could be represented as a triple KO - AV - CX, thus describing the ability of putting into action a given KO in a specific context.

Figure 1 shows the webpage for the collection/modification of professional figures/qualifications, together with the list of learning outcomes related to the *to prepare bar services* task, belonging to an Italian Bartender professional figure.

**B. Annotation of professional figures/qualifications**

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, is being 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 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.

**C. Testing**

In this phase, some migrant's résumés, companies requests and qualifications are inserted in the knowledge base. In order to simplify this task, a web page similar to the one of existing search engines is being developed. In this page the migrant (or company or training institute) 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 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 metabolized by an animal*, while, 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.

Figure 2 shows the webpage for the annotation of a learning outcome: here the user specified the ability to *utilize aliments' warehouse*, and the platform suggested him a ranked list of similar learning outcomes (on the left side of the page), with found concepts written in bold characters. The user could then select one or more learning outcomes to describe his/her



Figure 2. Annotation of the *utilize aliment's warehouse* learning outcome

characteristics (in this case, selected items will appear in the right side of the page).

The testing will be performed on job profiles or qualifications belonging to the sectors identified in the first stage, and will provide a measure of the percentage of correctly annotated learning outcomes as well as a degree of quality of matchmaking results. It is possible that for some knowledge, skill or competences the platform could not be able to find the corresponding pre-annotated learning outcome, or could not be able to compute a correct match: this result should not be considered as a failure of the methodology, but rather it should encourage to widen the WordNet database, since, probably, some relations among synsets could be missing.

#### *D. Development of additional features and dissemination*

This step will be carried out with the collaboration of local employment agencies, and will be devoted to the development of additional features that could be interesting for job seekers, companies and training institutes, such as the manual annotation of learning outcomes, a graphical representation of matchmaking results with respect to owned (or required or provided) learning outcomes, the possibility to receive suggestions on training courses to attend in order to acquire further competences, in the lifelong learning view, the possibility to create different résumés tailored to different job advertisement categories, etc. Additionally, a massive dissemination of the LO-MATCH platform will be carried out: at this purpose, a conference with Training Authorities representatives and stakeholders will be planned.

### V. CONCLUSIONS

In this work the LO-MATCH platform, a semantic tool designed to support the identification/development of required/missing competences currently under development in the framework of the MATCH project, together with the four steps of the methodology for its development, are presented. This tool relies on the semantic thesaurus WordNet in order to express résumés, companies' needs and qualifications according to a shared ontology. By the exploitation of the LO-MATCH platform, job seekers' abilities could be matched with companies' needs, while students and workers could receive suggestions on training courses to attend in order to improve their competences, in a lifelong learning perspective. Future steps will be devoted to testing the presented search engine on a set of job profiles/qualifications and developing additional features that could make the LO-MATCH platform even more usable.

### ACKNOWLEDGMENT

The work discussed in this paper has been supported by the MATCH "Informal and non-formal competences matching devise for migrants employability and active citizenship" project (<http://match.cpv.org>), ref. 510739-LLP-1-2010-1-IT-GRUNDTVIG-GMP. This project has been funded with support from the European Commission. This document reflects the views only of the author, and the Commission cannot be held responsible for any use which may be made of the information contained therein.

### REFERENCES

- [1] Europass, [http://europass.cedefop.europa.eu/europass/home/hornav/Introduction.csp?loc=en\\_GB](http://europass.cedefop.europa.eu/europass/home/hornav/Introduction.csp?loc=en_GB). Cited November 2011
- [2] "Recommendation of the European Parliament and of the Council of 23 April 2008 on the establishment of the European Qualifications Framework for lifelong learning" <http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:C:2008:111:0001:0007:EN:PDF>. Cited November 2011
- [3] "ECTS Users' Guide" [http://ec.europa.eu/education/lifelong-learning-policy/doc/ects/guide\\_en.pdf](http://ec.europa.eu/education/lifelong-learning-policy/doc/ects/guide_en.pdf). Cited November 2011
- [4] T.R. Gruber, "Toward principles for the design of ontologies used for knowledge sharing," Padua workshop on Formal Ontology, 1-23, 1993
- [5] WordNet <http://wordnet.princeton.edu/>. Cited November 2011
- [6] P. Poyry, K. Peltto-Aho, J. Puustjarvi, "The role of metadata in the CUBER system," Proceedings of the 2002 annual research conference of the South African institute of computer scientists and information technologists on Enablement through technology, 16-18, 2002
- [7] P. Poyry, J. Puustjarvi, "CUBER: a personalised curriculum builder," Proceedings of the 3rd IEEE International Conference on Advanced Learning Technologies, 326-327, 2003
- [8] P. Bourque, L. Buglione, A. Abran, A. April, "Bloom's Taxonomy Levels for Three Software Engineer Profiles," Eleventh Annual International Workshop on Software Technology and Engineering Practice (STEP 2003), 123-129, 2003
- [9] B.S. Bloom, "Taxonomy of educational objectives - the classification of educational goals - Handbook I: cognitive domain," Longman, 1956
- [10] M. Hoffmann, "Using Bloom's Taxonomy of learning to make engineering courses comparable," 19<sup>th</sup> EAEEE Annual Conference, 205-209, 2008
- [11] V. Gatteschi, F. Lamberti, A. Sanna and C. Demartini, "Using taxonomies and ontologies to define occupation and education-driven European qualifications," Int. Journal on Computer Science and Information Systems vol 6, 30-44, 2011
- [12] Hexin Lv, Bin Zhu, "Skill ontology-based semantic model and its matching algorithm," 7th International Conference on Computer-Aided Industrial Design and Conceptual Design CAIDCD '06, 1-4, 2006
- [13] B. Pernici, P. Locatelli, C. Marinoni, "The eCCO System: An eCompetence Management Tool Based on Semantic Networks," Workshop on Ontology Content and Evaluation in Enterprise (OnToContent 2006), 1088-1099, 2006