Using Tag Clouds to Support the Comparison of Qualifications, Résumés and Job Profiles

V. Gatteschi, F. Lamberti, A. Sanna and C. Demartini
Politecnico di Torino - Dipartimento di Automatica e Informatica, Torino, Italy
valentina.gatteschi@polito.it, fabrizio.lamberti@polito.it, andrea.sanna@polito.it, claudio.demartini@polito.it

***Abstract*—The labor market is today characterized by a marked competitiveness. The ability to accurately choose the right education and training paths leading to the acquisition of the right competences required for a given job position, as well as the capability to properly select human resources based on precise job requirements are becoming more and more key factors to success. Nevertheless, the comparison of education opportunities generally requires a manual analysis of huge qualification descriptions and course syllabi, whereas staff selection (or job seeking activities) often relies on non-automated processes mainly based on an extensive check of applicants' curriculum vitae (or of possible job offers). The above tasks are extremely time consuming and, given the large amount of information to be considered, they also risk to provide unsatisfactory results. In this paper, the above issues are addressed in the perspective of semantic technologies. In particular, a web-based application is presented, which exploits ontological descriptions and tag cloud-based visualization strategies to generate a direct representation of the overlap between learners' needs and existing education and training paths, as well as between job seekers' profiles and company requirements. This way, potential users are provided with an effective support for matching job and study offers with corresponding demands.**

I. Introduction

In recent years, learners' and workers' mobility became a relevant topic in the European legislation. In fact, in order to enhance the comparability of qualifications across Europe and increase the competitiveness of the European labor market, several initiatives, such as the Bologna Process [1] and the Bruges-Copenhagen Process [2] have been undertaken. Nevertheless, while in higher education mobility between Universities and recognition of prior learning are almost a praxis, a fully inclusive society where learners' and workers' learning outcomes could be accumulated in a comprehensive lifelong learning perspective is still under construction.

One of the main obstacles to the implementation of the above vision was represented by the shortage of tools capable of fully supporting the readability, transferability and comparability of qualifications. In 2008, the European Parliament and the Council took a first step to address such needs by establishing the European Qualification Framework (EQF) [3], a common reference system acting as a translation device to make qualifications readable and understandable across different European countries and systems.

According to the EQF guidelines, the above objectives could be reached by adopting a rigorous classification of all qualifications based on eight reference levels, and by precisely defining the semantics of associated learning outcomes (expressed in terms of knowledge, skills and competences), thus opening the way for the creation of a shared understanding in the lifelong learning domain.

However, the creation of a European-wide framework is only part of a more complex process: in fact, even though the EQF defines a shared format for cataloguing and expressing qualifications, concrete achievements of individuals (either resulting from formal, non-formal or informal education and training processes) have to be expressed with a syntax-independent formalism capable of overcoming linguistic and cultural (i.e., semantics) barriers.

In this paper we introduce the LO-MATCH platform (http://www.lo-match.polito.it), a web-based tool that is being implemented in the context of the MATCH "Informal and non-formal competences matching devise for migrants employability and active citizenship" project. The platform relies on semantic technologies to tackle heterogeneity issues in the descriptions of qualifications/résumés and labor market's needs due to the use of non-shared vocabularies. Moreover, it exploits a tag cloud-based visualization technique to quickly depict aspects to be considered in the mobility and job seeking phases. Specifically, tag cloud properties, like font size and distance from the center of the cloud, are used to provide an immediate overview of the main characteristics of a given qualification with respect to specific learner's needs, as well as to highlight key job seeker's attitudes with respect to a particular job offer (both from the job seeker's and the employer's points of view).

The rest of the paper is organized as follows. In Section II, research activities focused on the construction of tag clouds are presented, together with several application possibilities. Section III illustrates the main idea behind the designed platform, by discussing the steps required for creating the repository the overall approach is built upon, by analyzing the proposed tag cloud-based presentation strategy and by finally presenting the envisaged usage scenarios. Finally, Section IV provides conclusions and shows open research directions to be possibly investigated in the future.

II. Background

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

978-1-4577-0052-1/11/$26.00 ©2011 IEEE

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 [4] color information was included to visualize the actuality of tags. In [5], 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 [6], the constraint of rectangular layouts was removed, and a graph-based structure was used to visualize relations between tags. In [7], 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 [8], tag placement based on similarity was exploited, by clustering similar tags in the cloud based on co-occurrence. A different approach was taken in [9], 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 analyzed 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 [5][10].

Indeed, also according to the outcomes of the above studies, advancements on this topic will definitely benefit of practical study cases demonstrating the effectiveness of tag cloud-based techniques in concrete application scenarios like the one presented in this work.

III. Matchmaking and Tag Clouds

As said above, in order to help learners and job seekers in the identification of education/training or working opportunities better fulfilling their needs or expectations (in terms of missing or matching knowledge, skills and competences, respectively), as well as to support companies in the selection of the right candidates for a given job position, in the framework of the MATCH project the LO-MATCH semantic platform was designed.

The proposed web tool exploits an ontology, i.e., an *explicit specifications of a conceptualization*, as defined in [11], to describe qualifications/curriculum vitae and occupational profiles (expressing labor market's needs) collected in the LO-MATCH knowledge base. In order to make the above elements comparable, descriptions have been structured in terms of learning outcomes according to the EQF indications, and each learning outcome has been annotated (i.e. marked with one or more tags) in a manual or semi-automatic way by making reference to concepts defined in the ontology.

In the ontology, concepts are linked to other concepts by means of relations, which mostly belong to the subsumption category. Subsumption relations contribute to the creation of the overall hierarchy of concepts/tags (taxonomy) that allows the platform to deal with learning outcomes expressed at different levels of details, thus improving comparison results.

Collected information is then exploited to draw a cloud-based representation of a qualification (when the learning dimension is took into account), or of job applicant's characteristics/company's requirements (when the job seeking/hiring domains are considered). In the following, the methodology for the construction of the knowledge base, as well as the approach for the generation of the tag cloud-based representation are presented by making reference to specific usage scenarios.

A. Construction of the Knowledge Base

In order to ease the insertion of relevant information into the knowledge base, a semi-automatic mapping procedure has been created. In a nutshell, the functioning of such a tool could be summarized as follows: when the user specifies a new learning outcome (e.g., as part of a qualification/résumé or of an occupational profile), the system automatically detects and suggests him or her the relevant concepts that could be linked to each word in the newly introduced element.

In order to perform this task, the tool exploits the Wordnet repository [12], an English thesaurus collecting lexical and semantic relations among terms. When the user chooses a particular concept for annotating a given word of the selected learning outcome, the lexical/ semantic relations in the ontology are also recorded (together with related concepts). When a particular word is not found in the repository, the user can specify another term he or she may consider related somehow to the initial one (e.g., it could be more generic, more specific, it may share the same definition, etc.). Concepts associated to the new term can then be used to annotate the selected learning outcome word. This way, the initial scope of the ontology is extended, and new annotations could possibly rely on a more complete set of concepts and relations.

For each concept linked to a given learning outcome, a value of mastery/importance has to be provided (when the user considers some concepts more important than others). The difference between degree of mastery and importance is linked to the particular kind of end user working on the platform: in fact, when an education and training actor (or a job applicant) inserts a qualification (or curriculum vitae) in the knowledge base, he or she has to specify a degree of mastery, whereas when a company (or a learner determined to find a training path fulfilling his needs) is inserting its requirements, it has to specify a degree of importance. In other words, the degree of mastery refers to the (education/training or job) offer perspective of the matchmaking process, whereas the degree of importance is related to the demand side.

As a matter of example, Fig. 1 shows the graphical interface that allows companies and learners to specify

Figure 1. Insertion of a new concept in the set of company's/learner's requirement.

new requirements together with the associated degree of importance. In particular, the degree of importance could assume the following values: *low*, *medium-low*, *medium*, *medium-high*, *high*.

B. *Generation of the Tag Cloud-based Representation*

Concepts stored in the knowledge base and their degree of importance/mastery are used to draw a cloud-based representation of a) the characteristics of the qualification better satisfying learner's requirements, b) the features of a job seeker's curriculum vitae better matching company's requirements, c) 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., learner's requirements and learning outcomes associated with a particular qualification, company's needs and job seeker's characteristics, etc. Thus, even non-skilled users/operators could easily see why a given matching has been obtained.

TABLE I.
DEGREE OF MASTERY FOR KNOWLEDGE ELEMENTS EXPRESSED BY TWO JOB SEEKERS APPLYING FOR A SHOP ASSISTANT POSITION AND IMPORTANCE IN THE COMPANY'S PERSPECTIVE

Knowledge element (concept)	First applicant	Second 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

When focusing on the point of view of a learner looking for a qualification capable of filling his or her learning outcome gaps (or on the perspective of an employer searching a worker to hire), the font size used for drawing the tags is determined by sorting learner's (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 $x = r\cos(\theta)$ and $y = r\sin(\theta)$. In such expressions, r is defined as $R(1 - m + D)/D$, 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, and θ is a random angle. More details on the above steps are reported in [13].

The toy example reported in Table 1, 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 $x = 181$ and $y = -84$ (assuming the center of the cloud in $x = 0$ and $y = 0$).

Fig. 2 and Fig. 3 show the tag clouds for the curricula of the two applicants, based on the taxonomy reported in Fig. 4: 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 (Fig. 2) 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,

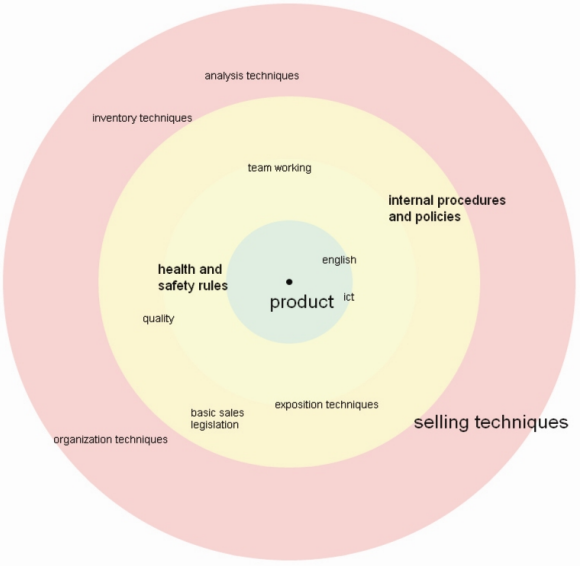

Figure 2. Tag cloud-based representation of the first applicant's curriculum vitae in the company's perspective.

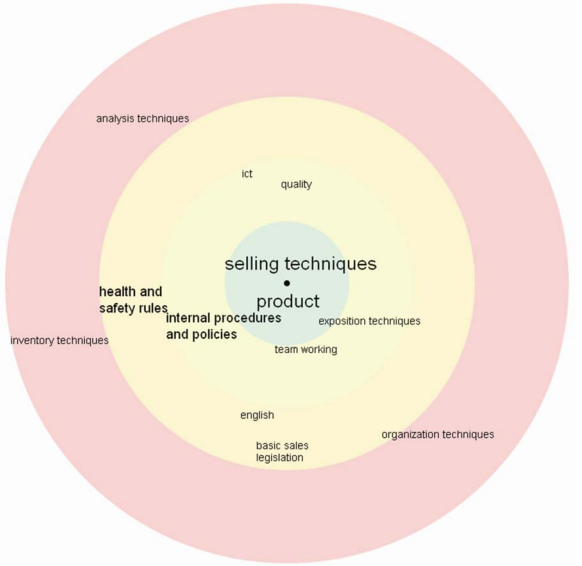

Figure 3. Tag cloud-based representation of the second applicant's curriculum vitae in the company's perspective.

like *selling techniques*, are placed on the external area (thus underlying their lack). In turn, the second applicant (Fig. 3) 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.

The above examples analyze 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

Figure 4. Portion of the ontology of interest for the tag clouds exemplified in Fig. 2 and Fig. 3.

companies that could recognize their abilities.

The interface designed to this purpose is depicted in Fig. 5 (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 *organization techniques*, *inventory techniques* and *analysis techniques*.

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. It is worth remarking that, even though for sake of simplicity the examples above only dealt with knowledge elements, in the MATCH project the described approach is actually used to draft a comprehensive tag cloud-based representation of EQF compliant learning outcomes, with knowledge elements linked to action verbs and to context information in order to express skills and competences, respectively (according to [14]).

IV. Conclusion and Future Works

In this paper, a tag cloud-based application supporting qualifications comparison and job matchmaking is presented. The proposed tool relies on a knowledge base containing qualifications, curriculum vitae and job profiles expressed in terms of knowledge, skills and competences. Information stored in the knowledge base have been annotated by exploiting an ontology initially based on the

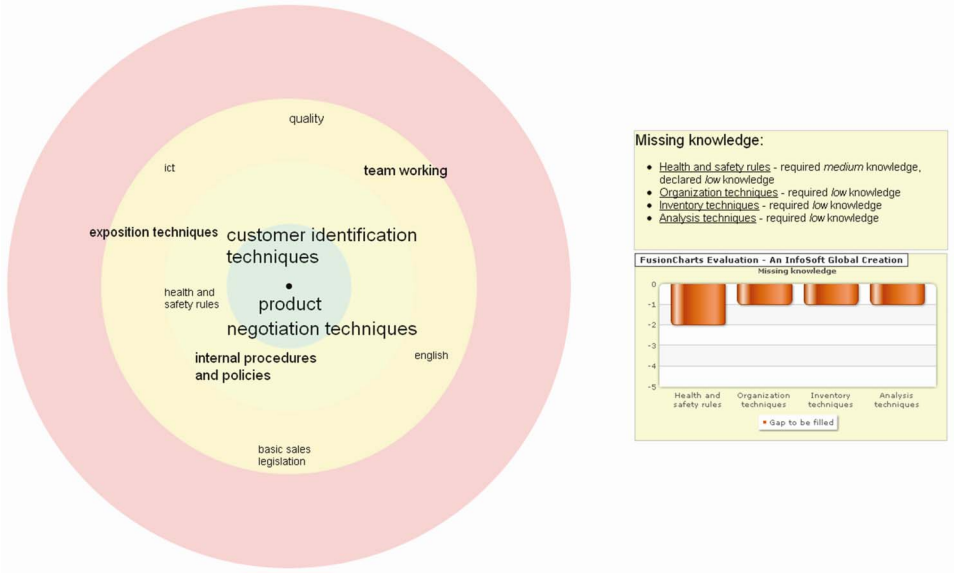

Figure 5. Tag cloud letting the second applicant (whose knowledge is reported in Table 1) compare his or her expertise with company's requirements.

Wordnet database, which has been later extended by the users, where needed. Target users could be either learners, job seekers or companies. In fact, learners could exploit the platform to easily and effectively find qualifications fulfilling their education and training needs. Job seekers could use the tool to find the job offers better matching their abilities. Finally, through the platform, companies could get an immediate overview of the expertise of candidates applying for a given job position. On the one hand, thanks to the use of a uniform notation for describing aspects which may be expressed in different terms and at different levels of details by the various actors, the proposed tool aims at overcoming lexical and semantic barriers between education, training and working offers and demands. Moreover, by exploiting trivial properties of tag cloud-based representations (like font size and distance of tags from the center of the cloud), an immediate overview of aspects of interest resulting from the comparison can be provided to heterogeneous users.

Future works will be devoted, on the one side, at considering additional factors in the construction of the tag cloud-based representation. On the other side, efforts will be devoted to extend available features, e.g., by introducing the possibility for the learners to identify training courses providing learning outcomes associated with a particular qualification or by letting the platform suggest competences and learning modules to be considered by the companies for on-the-job continuous training. Finally, the presented methodology will be extended to other sectors, and subjective tests will be performed with the aim at evaluating the usability and acceptability of the designed platform as well as to quantify its added value with respect to other tools for comparison and matchmaking like, for instance, text-based facilitators.

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

REFERENCES

[1] "Bologna Declaration Communiqué of the Meeting of European Ministers in charge of Higher Education (19 May 2001)".

[2] "Declaration of the European Ministers of Vocational Education and Training, and the European Commission, Convened in Copenhagen on 29 and 30 November 2002, on Enhanced European Cooperation in Vocational Education and Training - The Copenhagen Declaration".

[3] "Recommendation of the European Parliament and of the Council of 23 April 2008 on the Establishment of the European Qualifications Framework for Lifelong Learning".

[4] 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.

[5] 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.

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

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

[8] 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.

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

[10] 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.

[11] T.R. Gruber, "Toward Principles for the Design of Ontologies Used for Knowledge Sharing," *Int'l J. Human-Computer Studies*, vol. 43, no. 5/6, pp. 907–928, 1995.

[12] C. Fellbaum, "WordNet: An Electronic Lexical Database". MIT Press, 1998.

[13] V. Gatteschi, F. Lamberti, A. Sanna, C. Demartini, "A Semantic Matchmaking System for Job Recruitment", in *Proc. of the 10th International Conference on Knowledge Management and Knowledge Technologies*, pp. 50–59, 2010.

[14] B. Pernici, P. Locatelli, C. Marinoni, "The eCCO System: An eCompetence Management Tool Based on Semantic Networks," in *Proc. of the Workshop on Ontology Content and Evaluation in Enterprise*, pp. 1088–1099, 2006.