

Defining a Novel Ontology for Educational Counselling based on Professional Indicators

Antonio Robles-Gómez¹, Salvador Ros¹, Antonio Martínez-Gómez², Roberto Hernández¹, Agustín C. Caminero¹, Llanos Tobarra¹, Rafael Pastor¹, and Jesús Cano¹

¹ Control and Communications System Department, ETSI Informática
Universidad Nacional de Educación a Distancia (UNED), Spain
{arobles, sros, roberto, accaminero, llanos, rpastor, jcano}@scc.uned.es

² Student, ETSI Informática
Universidad Nacional de Educación a Distancia (UNED), Spain
martinezgamez.antonio@gmail.com

Abstract. This paper proposes a novel ontology to adjust or associate personal studies with training and professional profiles, in order to classify the relevant information related to job market. This way, educators will be able to help students to improve their counselling capacities in a personal way. Prior to the ontology definition, two steps are required: 1) obtaining data from professional networks by using scraping techniques; and 2) analyzing it to select a set of relevant parameters (or indicators). These indicators are used for building a standardized ontology, which establish relationships among educational profiles and professional careers, so making students' recommendations about employability needs.

Keywords: Learning Analytics (LA); Education; Professional Career, Indicators, Web Scraping.

1 Introduction

A large amount of data from various information sources on Online Social Networks (OSNs), and the related information of a great number of potential users in them, can be nowadays employed for different purposes. Therefore, the use of massive processing techniques and algorithms becomes necessary in order to integrate and analyze such information. In our case, we focus on the field of education, where the modeling of training and professional profiles will be defined by a novel ontology. This ontology will also be employed to make students recommendations about employability needs. As a clear and simple example of it, imagine that a user have a particular degree, but most of his/her jobs of interest require a higher degree, we can recommend him/her to improve her/his educational profile, for instance by starting a particular post-degree.

In this sense, learning does not take place in an isolated context, but occurs in a social, cultural and working environment, which to some extent determines decisions made by students and administrations, guiding their careers. All these dimensions are

sometimes complex to collect and limited to the internal data generated at University. It is therefore important to add/match these items to a complete analysis (including both the educational and employment contexts). The difference between academic preparation and labor market needs is not a new issue, among others [1] addressed it from a traditional point of view.

Therefore, the study of external sources of information oriented to professional market (such as InfoJobs [2], Monster [3], or TalentJob [4]), as it is done in this work, includes educational requirements, among others. The use of social networking in education as a motivating tool and its use to determine the impact of education on labor market integration is not a new issue. For instance, LinkedIn is employed in [5]. Once the gathered information is analyzed, a set of relevant parameters (or indicators) are selected to be used in our ontology. This novel ontology adjusts personal studies with training and professional profiles, in order to classify the job offers. As a consequence, educators will be able to help students to improve their counselling capacities in a personal way. We can also make professional recommendations from the users' educational profiles.

The structure of this paper is as follows: Section II describes how the selected information is gathered. After that, Section III analyzes the data obtained by selecting a set of professional indicators. The proposed ontology is presented in Section IV. Finally, Section V discusses our conclusions and suggests guidelines for future work.

2 Obtaining Data from OSNs

The information hosted on the web is normally unstructured, or at most semi-structured –in the case that access mechanisms have been provided to developers. Some websites take advantage of APIs (Access Programming Interfaces). These APIs operate as a bridge between the users' browser and the REST services and they allow getting JSON responses to source's invocations by using the field selectors [6]. In this sense, and according to the user's permission grants, several professional OSNs allow getting information, such as the actual job position, job requirements, summary description, and so on. However, this method usually has restrictions on the permission access. For instance, the LinkedIn API of is very limited due to the great restrictions. For this reason, this data collection method is not the most appropriate for our purposes, as already stated in [7].

As a consequence, a scraping strategy has been elaborated for this work, in order to get as much information as possible [8] [9]. In this sense, we have applied a crawling technique using the Scrapy framework on Python 2.7. Our technique relies on HTTP requests following the creation date in order to explore the different levels starting from the seed and by reaching leafs which represent public profiles. Other exploration methods could have been employed. During the exploitation phase, we dealt with regular expressions corresponding to generic XPath's that look into HTML code standing for each public profile and extract required items. A prototype of the crawling system has been developed, and graphical interfaces are being incorporated to the system at the moment.

The sources of information used for this work have been InfoJobs and Monster professional OSNs. We are also currently gathering this information from additional

professional OSNs, such as LinkedIn [7]. Fig. 1 shows an example of a “job offers” search in InfoJobs. This image is taken from <http://www.infojobs.net/>, where personal offer details have been removed for the sake of privacy. The most interesting thing of it is that each offer is composed of a set of parameters, such as location, category, minimum degree, years of experience, and so on. The major problem is that the organization of information is not clear and depends on the particular OSN.

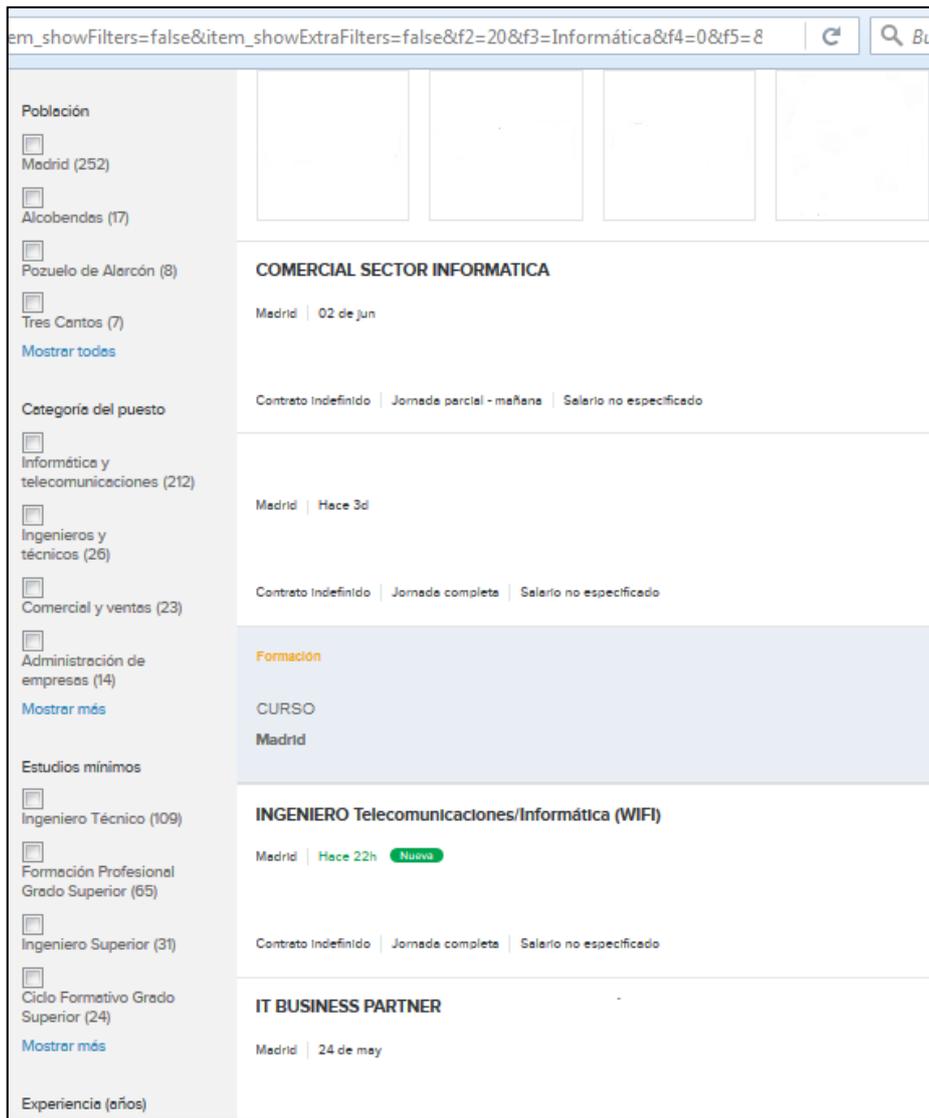


Fig. 1. Example of a “job offers” search in the InfoJobs OSN (image taken from <http://www.infojobs.net/>, where personal offer details have been removed)

3 Analyzing Data from OSNs

In our particular case, after obtaining a dataset, we analyze the gathered information by applying some filters for establishing and standardizing these indicators, as a previous step for building our ontology, which deals with special features of professional networks. All datasets are dynamically updated in our scraping system to keep past and current offers, and the obsolete offer positions will be moved to a secondary database.

Table 1 shows the selected indicators (or entities), which will compose the ontology. The first indicator which is relevant for our purposes is the short description of the job offer, named *Title*. Related to this indicator, we have other two parameters: the long description of the offer (*Description*) and the period of the offer (*Duration*). We have categorized these indicators as *Principal* and *Secondary*, to indicate their relevance in some way, since all of them can become important.

In addition to this, the *Category* and *Sub-Category* are gathered from the job offers, these one are composed by a set of nominal values (already established by the websites dedicated to professional purposes). The InfoJobs values for these indicators are taken as a basis, because of the fact they are more completed.

Then, we have the *Company* and *Company Type* indicators. These reflect the company name and type of job of the offer, respectively. The *Location* of the job position is also obtained, in terms of *City* and *Province*. Once this value is filtered, some students could only revise offers in his/her location.

Furthermore, the *Requirements* are also obtained. This field contains *Experience*, *Degree Title*, *Degree Title Type*, and *Other Requirements*. We highlight that the same degree's title may be used in different countries for representing different educational levels. For this reason, a semi-automated categorization of the degree's titles will be employed in our system. The UNESCO's ICSED [10] education degree levels classification presents a revision of the ISCED 1997 levels of education classification. This point can connect with the education purposes of users, in order to improve their professional profiles.

Additionally, the *Salary* details are obtained. From the filtering of these fields, we obtain the *Min-Salary*, *Max-Salary*, *Currency* and *Frequency* values. These indicators could be relevant as a refinement of this study.

Finally, some details about the source of information are given, that is, the *Source* (InfoJobs, Monster...) and the *Creation Date* of the offer. These last indicators are useful to remove obsolete offers, among other tasks. Although it is not included in the ontology, the URL and offer code is also maintained in the database for administration purposes.

Some indicators will be common to indicators proposed in [7] for educational profiles within the academic analytics context. We are currently extending their students' educational profiles to be matched with selected indicators, so making students' recommendations about the employability needs.

Table 1. Selected indicators for the web-semantic ontology

Indicator	Level	Description	Example
<i>Title</i>	Principal	Short description of the job position	Java Programmer
<i>Description</i>	Secondary	Details of the job position	This position focuses on programming physical devices
<i>Duration</i>	Secondary	Period of the job position	2 years
<i>Category</i>	Principal	Category of the job position	Computers and telecommunications
<i>Sub-Category</i>	Secondary	Sub-category of the job position	Programming
<i>Company</i>	Principal	Company of the job position	Intel
<i>Company Type</i>	Secondary	Type of company of the job position	Other activities
<i>Location</i>	Principal	It is composed by city and province of the job position	-
<i>City</i>	Secondary	City of the job position	Madrid
<i>Province</i>	Secondary	Province of the job position	Madrid
<i>Requirements</i>	Principal	It is composed by the child indicators: Experience, Degree Title, Degree title Type, and Other requirements	-
<i>Experience</i>	Secondary	Minimum experience of the job position	3 years
<i>Degree Title</i>	Secondary	Minimum studies of the job position	Master
<i>Degree Title Type</i>	Secondary	Type of qualification of the job position	Computer Science
<i>Other Requirements</i>	Secondary	Additional requirements of the job position	B2 English
<i>Salary</i>	Principal	It is composed by the child indicators: Min-Salary, Max-Salary, Currency, and Frequency	-
<i>Min-Salary</i>	Secondary	Maximum salary of the job position	24.000
<i>Max-Salary</i>	Secondary	Minimum salary of the job position	27.000
<i>Currency</i>	Secondary	Currency for the salary of the job position	Euro
<i>Frequency</i>	Secondary	Frequency for the salary of the job position	Year
<i>Source</i>	Principal	Professional network of job position posted	InfoJobs
<i>Creation Date</i>	Secondary	Posting date of the job position	2016-06-12

4 The Ontology for Improving Students' Employability

In the computer science context, a new ontology is a definition of types, properties, and relationships among entities for a particular context of application. The ontology selects and categorizes the needed indicators, and it establishes their relationships [11] [12]. For the fields of artificial intelligence, web semantic, systems engineering, software engineering, and so on, it is usual to create ontologies to limit the complexity and organize information in a correct way [13], and solving problems on the particular context of application.

Fig. 2 shows the proposed ontology by using the indicators provided by the scraping techniques for the several OSNs studied at the moment, and organizing them to make students' professional recommendations depending on their educational profile. It follows the guidelines provided by the VIVO ontology [14], which is applied to a different context, but with a similar philosophy.

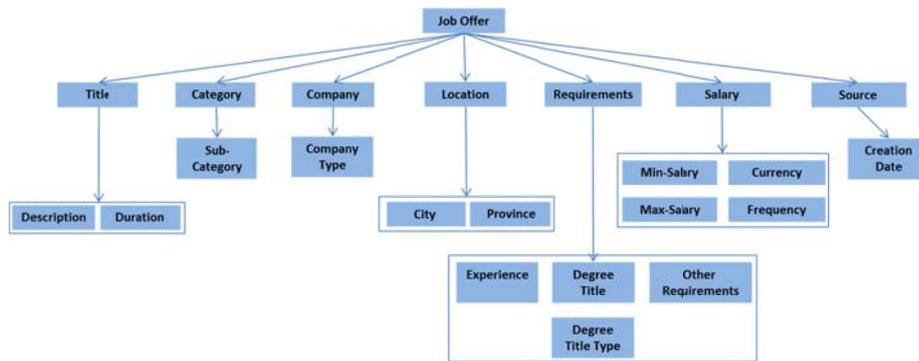


Fig. 2. Proposed web-semantic ontology for educational/professional profiles

A job offer is composed by a set of principal and secondary indicators, already defined and described above. The principal ones are: *Title*, *Category*, *Company*, *Location*, *Requirements*, *Salary* and *Source*. All secondary indicators are related to a principal one:

- *Title*. If the short description of the job position is not enough to make a recommendation, we can use the *Description* and the *Duration* of the offer.
- *Category*. If the category of the job offer is not enough to make a recommendation, we can use the *Sub-Category* value.
- *Company*. If the company of the job offer is not enough to make a recommendation, we can use the *Company Type* value. This will help user to know more details about what the company does for a specific offer.
- *Location*. This indicator is composed by *City* and *Province*, which is divided into these two specific indicators, to make easier the recommendations about locations.

- *Requirements*. The related secondary indicator is the most relevant ones for this study. It contains the required *Degree Title* and *Degree Title Type*, the *Experience* in terms of years, and *Other Requirements*. These indicators can easily match with the most relevant educational indicators for the student's profile.
- *Salary*. This indicator is composed by *Min-Salary*, *Max-salary*, *Currency* and *Frequency*, which is divided into these four specific indicators, to make easier the recommendations about salary; the salary range, and currency and frequency of payment.
- *Source*. This indicator contains the source of information for a particular offer, and it is directly related to the *Creation Date* of the offers. These indicators are thought for administration purposes.

Parts of the most relevant indicators that compose a educational profiles are contained in the proposed ontology. Our current educational profile includes the last and previous Degree Titles and Degree Title Types, among others. The location is also stored in the database. In this sense, we can recommend students a set of offers according to their degrees and locations. It will also be possible to recommend them additional studies if they can reach a set of offers with a higher position, more salary, or a more prestigious company, among others. We plan to make a refinement of the students' educational profiles in a nearby future.

For the testing phase of this work, invented information is being used. In case using real information of the University in the future, students' agreements will employed for achieving ethical issues. Anonymization techniques are already in use for privacy.

5 Conclusions and Future Work

Learning does not take place in an isolated context, but occurs in a social, cultural and working atmosphere. For this reason, it is very relevant to study and aggregate external information with current professional OSNs offers to users. This will help students to guide their education and, as a consequence, their professional careers. According to this, we have studied several professional sources of information from on the Internet, such as InfoJobs and Monster. From this reason, a large amount of data is currently been scrapped from these websites. Our system stores all the past and current offer positions. The time period of downloading each offer can be dynamically updated in real time.

This work also proposes a novel ontology to match personal studies with training and professional profiles, by selecting a set of relevant indicators from the information gathered from the professional OSNs. This ontology will establish relationships among educational profiles and professional careers, so improving the professional counselling by making students recommendations. Some guidelines from others ontologies, as the proposed in the VIVO project, have been used in this study as a reference.

On the other hand, we are developing a web application in order to recommend job offers to students from UNED University, by using and refining the matching professional and educational indicators. This system already implements the professional

ontology proposed in this work. Therefore, the recommendation step matches the selected indicators from our ontology with the educational profiles of students – an extension of the defined by [7] is being implemented. Students will additionally be able to perform searches according to certain criteria and offer position features, such as location, type of company, duration, and so on.

Acknowledgements. Authors would like to acknowledge the support of the research project (2014I/PPRO/031) from UNED and Banco Santander, and a local project (2014-027-UNED-PROY). Furthermore, we thank both the Region of Madrid for the support of E-Madrid Network of Excellence (S2013-ICE2715) and the Spanish Government for the support of SNOLA Network of Excellence (TIN2015-71669-REDT).

References

1. Trauth, E. M., Farwell, D. W., Lee, D.: The IS expectation gap: Industry expectations versus academic preparation. *MIS Quarterly*, vol. 17, no. 3, pp. 293–307, 1993
2. Infojobs. <https://www.infojobs.net/> (*accessed 2016-06-13*)
3. Monster. <http://www.monster.es/> (*accessed 2016-06-13*)
4. TalentJob. <http://www.jobandtalent.com/es> (*accessed 2016-06-13*)
5. Case, T., Gardiner, A., Rutner, P., Dyer, J.: A LinkedIn Analysis of Career Paths of Information Systems Alumni. *Journal of the Southern Association for Information Systems*, vol. 1, no. 1, Winter 2013
6. LinkedIn API. <https://developer.linkedin.com/> (*accessed 2016-06-13*)
7. Dai, K., Nespereira, C. G., Vilas, A. F., Redondo, R. P. D.: Scraping and Clustering Techniques for the Characterization of LinkedIn Profiles. In *Proc. 4th International Conference on Information Technology Convergence & Services*, pp. 1–15, January 2015
8. Liu, Y., Agah, A.: Topical crawling on the web through local site-searches. *Journal of Web Engineering*, vol. 12, no. 3, pp. 203–214, July 2013
9. Buccafurri, F., Lax, G., Nocera, A., Ursin, D.: Moving from social networks to social internetworking scenarios: The crawling perspective. *Information Science*, vol. 256, pp. 126–137, January 2014
10. Institute for Statistics of the United Nations Educational, Scientific and Cultural Organization (UNESCO): "International Standard Classification of Education: ISCED 2011". <http://www.uis.unesco.org/Education/Documents/isced-2011-en.pdf> (*accessed 2016-06-13*)
11. Al-Safadi, L. Abdulateef, N.: Educational Advertising Ontology: A Domain-Dependent Ontology for Semantic Advertising Networks. *Journal of Computer Sciences*, vol. 6, no. 9, 2010
12. Davies, J., Struder, R., Warren, P.: *Semantic web technology: trends and research in ontology-based systems*. Editorial: Wiley, April 2006
13. VIVO, ISF Ontology. <https://wiki.duraspace.org/display/VIVO/VIVO-ISF+Ontology> (*accessed 2016-06-13*)
14. VIVO Ontology Core.
<http://www.essepuntato.it/lode/owlapi/http://vivoweb.org/ontology/core/> (*accessed 2016-06-13*)