

A Life-long Learning Recommender System to Promote Employability

<https://doi.org/10.3991/ijet.v12i06.7166>

David Bañeres

Open University of Catalonia, Barcelona, Spain
dbaneres@uoc.edu

Jordi Conesa

Open University of Catalonia, Barcelona, Spain
jconesac@uoc.edu

Abstract—Is my professional knowledge outdated? Do I have the skills needed for the new challenges of the society? What knowledge do I lack to qualify for a job I like? What universities can I address to get knowledge that improves my employment expectations? These are relevant questions that all employees have done in any moment of their life. In addition, when there are high rates of unemployment and job offers that keep unfilled, the answers to these questions are even more relevant. Answering such questions open new opportunities for employed and unemployed people, by allowing them to design a formative plan according to their skills and expectations. It also provides evidences to employers about the skills and knowledge of the society, making them more aware of the skills of their potential future employees. The companies also will have more knowledge to design the professional career of their employees according to the company needs and the knowledge and skills of their employees. This paper proposes a system that helps people by showing which knowledge and skills a person misses for a given job position and what university courses the person can take to acquire the required skills and knowledge. The system has been implemented as a recommender system that helps users in planning their life-long learning. The paper shows the architecture of the proposed system, a case study to explain how it works, a survey to validate its usefulness and usability and some conclusions after its first experimentation.

Keywords—recommender system, employability, analytics, job finding, skills analysis, life-long learning

1 Introduction

Nowadays, organizations are facing new necessities when recruiting new employees. Advances in technology, availability of data, new society' challenges and new types of jobs require new skills that current staff or available people for hiring may not have. Such quick evolution has resulted in a very dynamic labor market. In that

sense, people may feel overwhelmed by the speed the labor market evolves and how quickly their knowledge and skills became obsolete. Therefore, it is relevant to provide analytical tools that support people to be aware of how well positioned they are for succeeding in their professional and job expectations, of what knowledge and skills they should get to be able to achieve such expectations or of how to maximize their employability.

That is not a necessity that impacts in persons (potentially employees) only, but also in organizations (potential employers). The necessity is even greater in environments where there are a high unemployment rate and environments where there are many job offers that become vacant because, supposedly, employees do not have the proper skills and knowledge. In the big data era, it is possible to create analytical tools that access and analyze data about job offers, about people resumes and about the academic programs that universities offer. These tools might provide relevant information to people in order to keep their professional knowledge updated, to increase their rate of success in further professional promotions and to raise their employability.

The objective of the presented work is to make a first step in the creation of such tools. In that direction, the present paper proposes a system that uses analytics and natural language processing techniques (NLP) to infer the knowledge a person should acquire to be able to perform a given job. The system is also able to identify the academic programs and subjects where the person can learn the missed knowledge and skills. Such information is used to provide a tool that helps users in improving their knowledge and skills according to their labor expectations. The proposed system uses NLP techniques to process the job offers automatically and extract the required skills and knowledge for each job offer. The skills and previous knowledge of users can be automatically inferred or manually provided by users. Then, analytics are used to create a recommender system that provides information about what knowledge and skills users should obtain and where they can be acquired.

The contribution of the proposal is twofold. First, it makes people more aware of their own limitations and more autonomous according to their further education, by providing tools to be self-aware of the obsolescence of their knowledge and to be able to plan their future formation more efficiently. Second, it can potentially reduce the gap between academic offer and labor market necessities, increasing the amount of people to be able to cover the job positions that actually keep deserted.

The paper is structured as follows. Section 2 presents a brief introduction to related work that analyzes the alignment of the academic offer and labor needs to increase the employability of population. It also includes information about standardization initiatives that have arisen to provide homogeneity in the curricula definition. Section 3 describes the proposed system by describing their main components, how the information is inferred and it is shown to users. Later, Section 4 shows, by a case study, how the system identifies the required skills for a given job position and the academic programs and courses a given user can take in order to get the lacking knowledge to apply to such offer. Finally, Section 5 concludes the paper by presenting some final remarks and pointing out further work.

2 Related Work

According to [1], cooperation between university and industry is one of the drivers of actual innovation. In addition, the participation of university students in the industry approaches them to their professional future. In that sense, it is important to take into account the industry to influence the development and evolution of professional academic programs in order to provide the skills and knowledge students will require for industry [2]. However, there is a misalignment among what is taught in academic programs and the real needs of the labor market [3][4]. Therefore, many people do not work in a job that corresponds to their knowledge and many employers consider that individuals do not have the knowledge and skills required to deal with the company challenges successfully.

The extraction of the professional knowledge of people and its alignment to the labor market has been tackled mainly by three approaches: services that look for alignment among curricula and labor market, services that look for the relation among education provided by universities and labor market and proposals done from a business intelligence perspective.

In the first direction, we have services such as Monster¹, StepStone², Lingway³ or Innovantage⁴ that propose job offers aligned to the person profile. Recommendations are made from a keyword analysis of the person resume and the job offers available. In some cases, the job offers are classified using ontologies, such as in [4][5][6][7]. Then, the comparison is more powerful, since it is not only performed with words but also with semantics, allowing to identify jobs that match a given resume even when there are not keywords in common between the resume and the job offer. However, these systems have some scalability problems for large and unstructured resumes, where it is difficult to infer all the keywords and their context. There are also other problems when using NLP techniques in order to extract the skills and content related to a person: same content and competencies can be written differently.

In the second direction, we have approaches focused on connecting an academic environment of a university with labor market [8]. In these systems, a web service that links academic programs to related jobs within web portals of job finding is provided to students. As far as we know, these systems use only the title of the academic programs to calculate the relationships, which can be conflictive since the title is very short, incomplete and do not describe well enough the content and practiced skills of a program or of a specific course.

From a business intelligence perspective, within the European Project *h-TechSight*⁵, the frequencies of key concepts of some technological domains were calculated. Their work provided detailed information about the demands of these domains geographically. Apart from the potential benefits for persons, such system may be useful for industry to be aware, in real time, the adoption and evolution of a

¹ <http://www.monster.com/geo/siteselection/>

² <http://www.stepstone.com/>

³ <http://www.lingway.com/>

⁴ <https://www.innovantage.co.uk/>

⁵ <http://banzai-deim.urv.net/~riano/hTechSight/projecte.html>

specific technology. However, under our knowledge, such system has not been applied to educational purposes.

Two common issues that appear in related work are the ambiguity of the definition of skills and knowledge within resumes and the interoperability of resumes and higher education programs. In order to deal with the ambiguity of knowledge and skills, there are some proposals focused on standardizing the resumes by providing a limited vocabulary for reducing term dispersion. Proposals for standardizing the skills and knowledge of persons are Europass⁶ and the e-Competence Framework⁷. Europass provides two documents to provide a standardization of resumes: the curriculum vitae, which presents the candidate skills and qualifications, and the European skills passport, which provides standardized information about the language skills of the candidates. It also defines the structure and content of another kind of official documents to document and share people competences in the European context. Examples of such documents are the *Europass mobility*, which records knowledge and skills that may be required by other countries, or the *Diploma supplement*, which records the knowledge and skills acquired by a given person in a higher education degree. Similarly, the *e-Competence Framework* provides standardization at skills level, but only for the area of Information and Communication Technology (ICT). To do so, it provides a common language and semantics for 40 competences, skills, knowledge and proficiency levels that can be understood across Europe.

Even when different institutions use the same specifications and standards to represent the information about their courses and the competencies the students get during their academic life, the interoperability continues to be an obstacle. Each institution and job portal will have their data isolated, generating different information silos. In order to surpass such problem, approaches used for other purposes should be adapted in a linked data global solution [9]. Some works that focus on the representation of information could be utilized in that direction. Authors in [10] use an ontology that represents learning resources to give support to online learning processes. In [11] it is provided an ontological version of the national UK curricula⁸ to represent the academic programs and subjects; and the *Learning in Process*⁹ project aims to empower people as their own mentors. To do so, they provide ontologies to represent the profile of students unambiguously and to allow knowledge exchange between different agents. Another work with an approximation similar to ours is presented in [12], which allows representing, among other information, the topics dealt in the materials of university courses. Since its focus is in the representation of knowledge, and not in their inference, the lists of the topics covered in each material and the material where they are covered should be specified by hand. The proposal is not focused to study how the academic program is aligned with the job necessities, but to evaluate the review of courses and see how different courses may overlap. Other approaches that use linked data to provide interoperability in the educational context can be found at [13].

⁶ <http://europass.cedefop.europa.eu/en/home>

⁷ <http://www.ecompetences.eu/>

⁸ <http://www.bbc.co.uk/ontologies/curriculum>

⁹ <http://www.calt.insead.fr/Project/lip/>

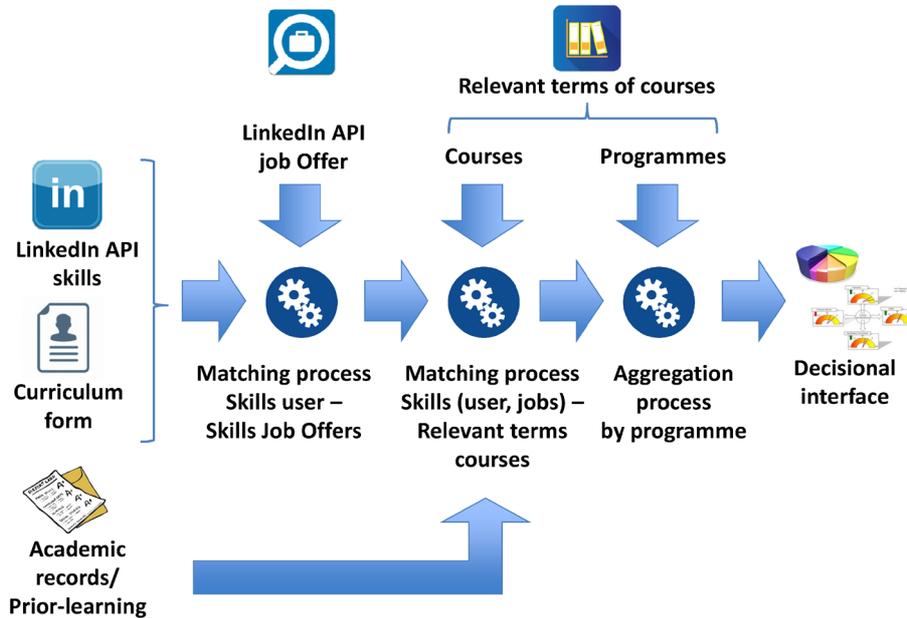


Fig. 1. Diagram of the program recommender system.

3 The Recommender System

In this section, the recommender system is described. As aforesaid, the system aims to recommend to each user what knowledge is required to apply to a specific job offer, the knowledge that the user is lacking and what are the best-suited academic programs to acquire such lacking knowledge. The required knowledge from job offers will be automatically inferred from the textual description of the job offer using NLP techniques. Therefore, the required skills (we will use the term skills to indicate both skills and knowledge in order to simplify readability from now on) are the real requirements that should be fulfilled to access the job and not the ideal skills commonly associated with a professional category.

The workflow of the proposed system is illustrated in Figure 1. As we can observe, the system has three different data inputs: information about the user profile, information about the job market, and information about academic programs.

First, the system needs all the information related to the prior-knowledge of the user to be aware what skills the user has. It is crucial to get the right and complete information about the user profile. The reason is that if a skill the user has is not introduced, then the system will fail to propose the right job offers to the user. The resume, job experience, skills, competences and even the prior-learning can be provided. For the resume information, the system provides two options: manual insertion using a structured resume form or an automatic extraction of the information stored in *LinkedIn profile*:

- Manual insertion of the profile: the system asks for information about language proficiency, previous and current job experience and skills. The user needs to add all this information systematically. The main drawback of this method is that the curriculum information has to be inserted using the custom format of the system, i.e. standard formats cannot be reused.
- Automatic insertion of the profile: the system uses the *LinkedIn API* to extract the skills and job experience of the user from its public profile. To do so, the *public Python LinkedIn interface library* (<http://ozgur.github.io/python-linkedin/>) has been used. This library is a Python interface that simplifies the interaction with the *LinkedIn API*. For instance, the skills extraction can be done using the simple method *get_profile* with the parameter *selectors=['skills']*. This method is more attractive than the previous one, since the user can reuse curriculum information inserted in LinkedIn. The drawback is that only this service is supported. However, the system has been designed to incorporate any public resume exporting service in the future easily.

Additionally, the system is capable of integrating the academic record of the user. Although this is not critical information for the alignment between resume and job required skills, this information is currently used to inform which courses are not required to be enrolled since they are already completed. As aforesaid, this functionality has been only implemented to deal with the academic record at the Open University of Catalonia. Other educational institutions may be integrated easily in the future since the system reads the academic record of the student in JSON format via a web service method. Thus, the endpoint has been already implemented in the service. Other universities will only need the specification of the endpoint to send the information of users in the JSON format.

The job offers are extracted from *LinkedIn* via the *Python LinkedIn library*, by using the method *search_job* with the parameter *selectors=['jobs']*, which returns a list of job offers. Additional search filters can be added to the method to improve the query and return an up-to-date list (i.e. the job offers in the last month).

For the academic information, we use the analytical data store provided in [15]. This data store contains an analysis of the most relevant terms associated with each course of the UOC that has been extracted by using similar techniques than the ones applied to job offers over the different materials of the subjects. The relevant terms can be any set of words (three words at maximum) that define skills, tools, competences or any other word associated with the knowledge of the course. Note that, this information has been extracted from the materials (textbooks) of the courses and manually validated by the instructors. Therefore, the relevant terms of each course describe the real knowledge of the course and not the ideal knowledge that should be learned based on teaching programs. Additionally, the data store is complemented with the structure of each program by giving the list of courses associated to each one. That allows getting the skills taught on each subject, but also on each academic program.

At this point the required data for the recommendation is gathered but it should be cleansed, summarized and improved using NLP techniques. These techniques are

applied to the information extracted from job offers and from student background (from the resumes and the courses the student has completed). The academic information does not require this kind of treatment, since the data has been already processed before being inserted in the analytical data store [15].

Once the data of the jobs offers and user background are gathered, their content should be analyzed to identify the skills contained within the text (ones will be the skills required for the job and the others the skills the user already has).

The skills extraction procedures use NLP techniques to easy manage and cleanse the data. The relevant skills are extracted by means of NLP techniques that segment the text, lemmatize the words and produce candidate terms, which are word lemmas, bi-grams and tri-grams of lemmas. Not all the candidates are relevant because of their part of speech (POS) and their importance in the text. So the extractor uses a POS tagger to analyze each text in order to identify the grammatical category of each term and lemmatize it. Then, the frequency of appearance is calculated to determine the importance of the candidate keywords in each text. Then, a cleansing procedure merges equivalent keywords and it discards stop words. The system uses Freeling [14] to determine when different keywords are different representations of the same lemma. If so, the different keywords are unified. For example, the keywords “implementing”, “implement” and “implemented” are semantically equivalent, therefore, they can be unified in one term “implement”. In the deletion of stop expressions, language-specific rules should be used to identify what expressions can be deleted. For example, a rule will be “When a bi- or a tri-gram contains a pronoun, then it can be deleted”. In addition, dealing with keywords of different length provides more terms and potentially more quality terms, but it can also produce some noise. In order to reduce such noise, a process to detect redundant keywords and to delete them has been created. Such process detects when there are two terms and one is a subset of the other. In such cases, we can discard the subsumed one when the frequencies of both terms are very close. For example, suppose the terms “analytics” with a frequency of 63 and “business analytics” with a frequency of 61. In such case the system can discard the term “analytics” without losing too much precision (only two apparitions of the term will be lost). Domain synonyms detection has not been dealt in the cleansing process since they require a high manual intervention.

Currently, extraction of users' skills and extraction of job offers use the same NLP approach. However, we are aware that they should be handled in a different way to improve quality extraction. Resume documents sometimes tend to use unstructured text to describe all the knowledge and experience. Similarly, job offers may also describe the required skills on unstructured text with another style. Developing specific NLP techniques depending on the discursive approach should improve the quality of the extracted information [[16]][[17]].

After gathering all these information, the system generates a dashboard with a list of reports. These reports show to each user which job offers are more related to the user capabilities and the list of academic programs and subjects that could help to acquire the missing skills to be fully eligible to apply for a job offer. The different phases the system follows to generate the final dashboard are (See Figure 1): skills

matching process, course matching process, and aggregation process. These phases are described in detail in the next subsections.

3.1 Skills matching process

The first module performs the matching process between the skills of the user and the skills required in the job offers. Matching skills are straightforward due to the parsing and linguistic cleansing that is done in the detection of keywords. In this stage, the system just looks for identical skills terms. The process computes two sets of terms:

- Acquired skills: job terms that describe the skills that have already been acquired by the user.
- Lacking skills: job skills not matched with the skills of the user. They are assumed as the skills the user lacks.

The acquired skills list is merely used to give further information to the user about its capabilities and chances for a given job position. The lacking skills list is useful to detect what academic programs offer support to their acquisition. Note that, user skills that are not necessary on the job (not matched) are discarded since they are not useful in further processes.

The process of analyzing all job offers generates a set of records of the form:

(user, job offer, score, list of job matched terms, list of job unmatched terms)

where the score provides information about how able is the user for applying to the job offer. It has been calculated as the number of matched terms between the resume of the user and the job offer.

This set of records is sorted by the score providing the list of the most relevant job offers related to skills of each user. This sorted list is transferred to the next process.

3.2 Course matching process

The second phase produces the matching between the selected terms in the previous step and the terms of the courses. As aforesaid, the analytical data store obtained from [15] provides information about the relevant terms for each course. The matching process is computed only between job unmatched terms and relevant terms of the courses since the matching process seeks the courses that cover the largest sets of unmatched skills for each job offer. The process generates records similar to the previous process but records are enriched with relevancy within the courses, where the score is defined as the number of matched terms between relevant terms of the course and job unmatched terms. This score gives information about the number of skills each course allows to acquire.

Additionally, this matching process is responsible for evaluating prior-learning of the users (regarding completed courses) and calculating the set of courses the user may potentially enroll to acquire the required skills. Since it is possible that the user has already completed such course, the system uses information from the academic

record of the user in order to see whether there is some selected course already done by the user. At this stage, the system cannot match completion of courses with prior-learning of the user out of the academic record. However, we are planning to use the technique described in [18] to support this feature.

This evaluation extends the record by adding information whether each of the selected courses has been completed by the user:

*(user, job offer, course, score,
list of course matched terms, list of course unmatched terms, is completed?)*

Like in the previous process, this set of records are sorted by the score, providing the list of courses that their learning contents are more related to the skills that the user needs to apply to the job offer.

3.3 Aggregation process.

The information gathered up until this stage is at the course level. In order to represent the information also at the academic program level, this process aggregates information by giving a complete vision of the best academic programs to improve user's skills.

The set of obtained records in the previous process is analyzed and combined using the academic program structure described in the analytical data store. Individual scores are summed on courses of the same program and, therefore, a new record is generated:

*(user, job offer, program, score,
list of program matched terms, list of program unmatched terms)*

All these records are stored in a data system to give full support to any decisional or recommender interface. Multiple reports can be provided at the user level, program level or even at the institutional level in terms of acquired skills or required skills to current job offers.

- At the user level, the system can provide fine-grained information related to which job opportunity the user is more eligible. This information may not seem valuable compared to the information provided by job search services. However, the value of the system is when this information is enriched with acquired and missing skills. Additionally, a user may be interested to see which programs and courses teach the missing skills, the relationship between courses and taught skills.
- At the program level, the produced information can be useful to see the relationship between programs and job offers. Potential students can be notified with information related to which career opportunities have a specific program; or the academic coordinator can even modify the content of the courses to get a better alignment with labor market and increase the employability of their students.
- At the institutional level, more deep analysis can be performed to get more quality curricula, detect open opportunities and improve the students' employability of the professional programs.

At this stage, we only provide support at the user level to give recommendations of the best programs to be enrolled to fulfill all requirements and skills to apply to a job offer. The next section describes a case study to see the output of the recommender system and an analysis of an opinion survey.

4 Experimental Evaluation

In this section, we present the experimental evaluation of the system. Two analysis are shown. First, a case study of the designed system to show the different reports the system offers. The case study has been done on a real student and the presented results are based on the current knowledge of the learner. Next, an opinion survey has been performed. The analysis has been extended to five students to analyze the potential of the system. We are aware of the limitation of the experimentation. However, the university employment service does not allow to perform large pilots on students on prototype systems. This first analysis will help to refine and improve the system and a large experiment will be done in the next step of this research.

4.1 Case study

In this section, we show the case study that is related to a real student enrolled in the Postgraduate of Business Intelligence at the Open University of Catalonia. The student has been selected since he has already done several courses of the master postgraduate and he has sought the guidance of the university employment service. We asked the student to be part of the experiment and he accepted. The student filled the resume based on the form provided by the tool since he did not have a LinkedIn account.

The system found different job offers based on information extracted from the *LinkedIn* job services. We performed a selection of job offers in the last month related to the professional category of *Information Technology* and at the living location of the student: *Barcelona, Spain*. Around 200 job offers were found related to the design of mobile applications, data analytics, web services expert and web design among others. The interface of the system filtered the results and showed only a list of 10 jobs based on the number of job matching skills. Note that, the user can modify the number of shown jobs but the default limit has been set to 10 based on the first experimentation. Note that, the correlation between jobs and curriculum was poor after the tenth job offer.

For this example, we found interesting to select a job offer not fully related to the background of the student since a job related to Business Intelligence would generate reports completely towards to the program of Business Intelligence. As we can observe on Figure 2, the selected job offer was *UX (User Experience) Designer*. The system provided a first bar chart visualization with the classification of the skills. As aforesaid, the skills of each job offer are naively extracted from the job services and, therefore, the system provides a large number of keywords. Some skills are closely related to the job, such as *ux*, and others are not relevant, such as *401k*. In the future,

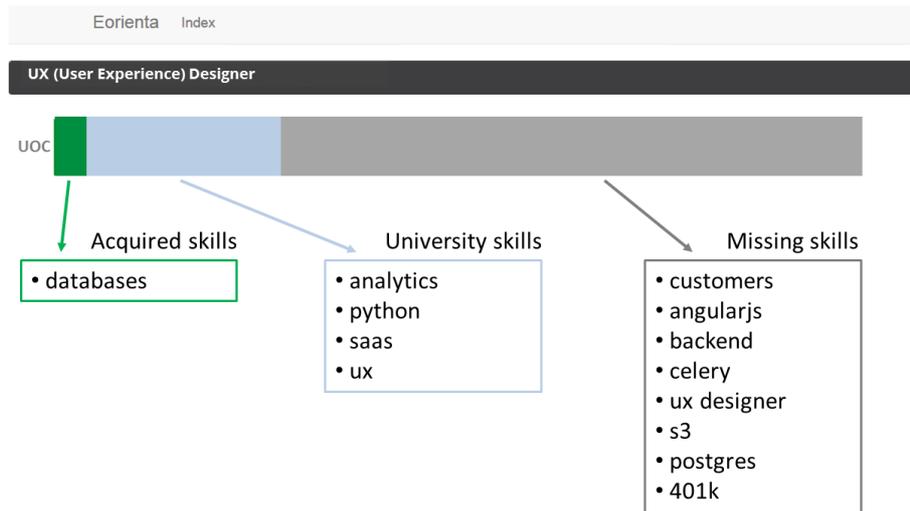


Fig. 2. Distribution of the skills of the job offer.

this extraction should be refined by reducing the outliers' keywords and finding equivalent keywords by using cleansing techniques [15]. The system classifies the skills in three classes:

- **Acquired skills:** These skills are the result of the matching process between skills, jobs and skills provided by the student in the resume form. This information is obtained from the skills matching process.
- **University skills:** These are the skills the user does not have, but that can be acquired from university courses. They are computed based on the union of the terms taught in all programs from the records generated by the second module course matching process. It is worth remarking that the skills can be learned from multiple programs (i.e. the skills are not exclusive of one program). At this level, it is not objective to analyze which academic programs and courses can be enrolled to acquire the skills; that will be the responsibility of the next report (see Figure 3).
- **Missing skills:** These are the missing skills that cannot be learned at the university. Mainly, in the example, because not all the programs have been taken into account in phase 2 of the system (i.e. not all computer science programs were dealt and therefore *angularjs*, *backend* and *postgres* were not matched with any course).

Currently, this report only offers one chart related to the knowledge learned in the Open University of Catalonia. However, the system is prepared to show reports of different universities when complete information about courses and programs is stored in the system. This comparison could help the learner to select the institution that could help to promote better his resume and knowledge.

When an institution is selected, more fine-grained information related to its programs is provided. The report (see Figure 3) shows the list of programs associated with the university skills. For each program, the list of the courses related to each skill

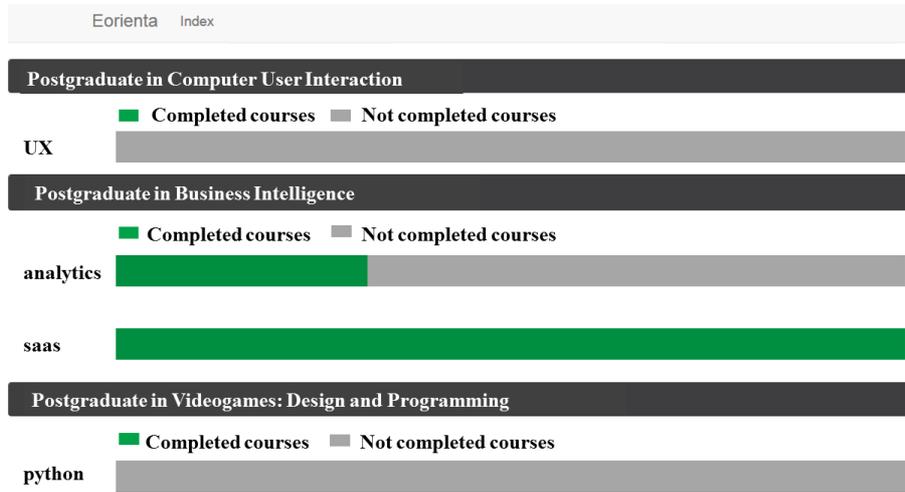


Fig. 3. Programs associated to the missing skills.

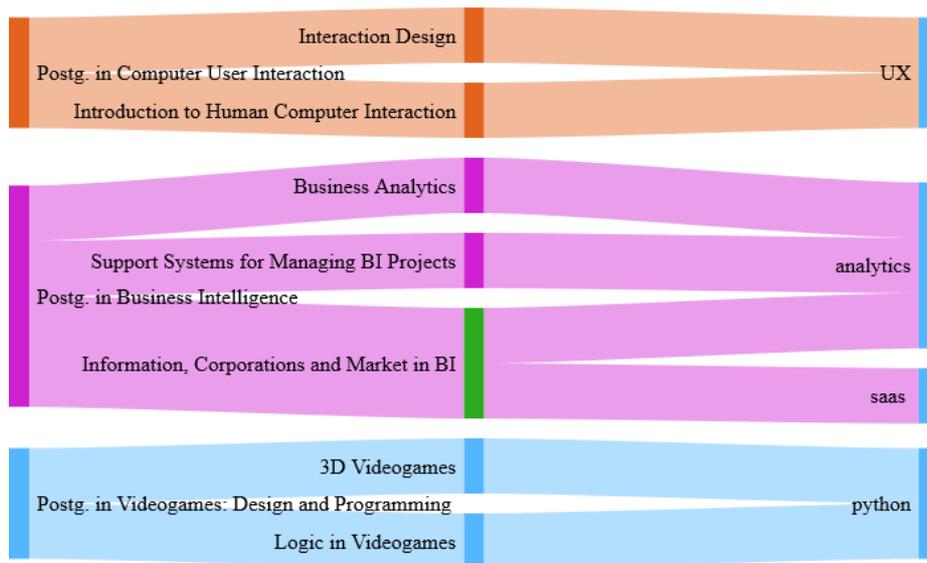


Fig. 4. Sankey diagram between programs and skills.

is also summarized. Contextualized information on each skill bar chart shows a detailed list of courses. Additionally, this report shows completed courses based on the academic record of the student in the university.

Finally, the last report is provided. A Sankey diagram (see Figure 4) shows the relationship between programs and skills. This diagram is useful to see all relationships between skills and programs, to see whether one skill is dealt with different subjects

(such as in the *saas* case) and to analyze which courses are involved in this relationship. Additionally, this report also shows completed courses with a different color to see a global view of completed courses. For instance, in the figure we can see that *analytics* is related to three courses *Business Analytics*, *Support systems for managing BI projects* and *Information, corporations and market in BI*. We can observe that the course *Information, corporations and market in BI* has already been completed by the student. Note that, the last one is also related to the *saas* skill.

We consider that these three reports provide sufficient information to users to freely select the job offer they want to apply, the list of institutions best-suited for improving their skills and the list of programs related to the missing skills. Although simple, we believe that the current system provides relevant information for a student to perform a selection.

4.2 Opinion survey

The second analysis is based on an opinion survey. The university employment service selected five students (including the previous one) who were seeking assistance for searching a job. Similar to the previous study, we requested their consent to be included in the pilot.

Students went through all the process: they filled the resume form, the system processed their information and a complete report was provided based on their curriculum and job preferences. When all the guidance process was completed, the survey was sent to evaluate the recommender system.

Five statements were requested to evaluate the opinion with respect to the new service. The Likert scale was used ranging from 1 to 5, being the larger value five meaning that the student strongly agrees with the statement. At the end of the survey, there is also an open question to gather critical comments and helpful opinions not collected in the previous statements. The statements are next described:

1. I think that the course recommender system is a valuable resource to improve my curriculum.
2. There is few manual intervention of the student.
3. The distribution of skills is easy to understand.
4. The skills distribution by program is easy to understand.
5. The graphical information about programs, courses and skills is easy to understand.

Table 1. Results of the opinion survey (1 to 5 scale, 5 is better)

Statement	Student					Avg.
	1	2	3	4	5	
1	5	5	5	3	5	4,6
2	4	3	3	1	3	2,8
3	4	5	3	2	4	3,6
4	5	5	5	4	4	4,6
5	5	4	5	3	5	4,4

The results of the survey are summarized in Table 1 where the average value for each statement is also reported. The first statement evaluates the global appraisal of the service. The objective is to know the opinion of this service to give a personalized support to students. We can observe that most of the students gave the maximum value. In the open comments, the student who valued the first statement with a 3, claimed that the recommender is not useful at all if the university does not give any additional support to find a job. We guess that the student wanted personalized guidance to get a job and not to continue studying to get better skills.

The second statement got the worse appraisal. The objective was to evaluate the consumed time by the students to use the system but all students focused the statement on evaluating the resume form. All students claimed that the resume form should be removed since they had already filled another resume for the unemployment service. One student stated that some standard curriculum form should be used instead and his opinion is that LinkedIn is not a standard curriculum form. Moreover, we can observe that no student complained about the different reports. Then, we assume that they did not require much time to analyze the results but alternative ways of getting the students' resume should be proposed.

Mixed opinions were found in statement 3. The objective of the statement was to evaluate the first report based on skills coverage by the university. The majority of the students gave a good value but one student complained about the report with a value of 2 out of 5. The student stated to be confused about the name of the skills. The student is enrolled in the Master degree of Security and he found skills such as *cloud*, *scrum*, or *positive*. As we can observe, a more deep cleansing process is needed to reduce outliers' keywords. We guess that, in some students, this type of keywords reduced the trustworthiness with the system.

The fourth statement aimed to evaluate the second report. In this case, a good result is also obtained with no comments. The report is easy to use and understand. Similar results have been achieved for the fifth statement. Students confirmed that the Sankey diagram is easy to understand. A comment was collected referring to the comparison of the different programs. The student claimed that the diagram should contain information about the minimal courses to be enrolled to obtain the maximum profit (i.e. maximum number of skills).

5 Conclusions and Future Work

In this paper, we presented a recommender system that allows people to be more aware of their own limitations and opportunities of employability. In addition, the system gives information to users that can be used to know what skills and knowledge they miss and how to acquire them. Therefore, the system could be a tool that supports users to self-design their long-life learning path according to job and professional expectations. The system is in a preliminary stage, but we believe it has large potential and may inspire new relevant and useful services, such as employment service, smart enrollment service, benchmark service of academic institutions, among others.

The case study presented in the previous section shows an excerpt of the analytical reports that can be provided at the user level. Additional services could be added at

the user level, but also reports can be provided at different levels (program, institution, geographical region, ...) as it is planned as future work. Other future steps are related to improving the quantity of data of the system, by analyzing new sources of information related to job offers and academic institutions, and the quality of the analysis, by using new NLP techniques particularly efficient when dealing with skills extraction. We are aware that the case study and the opinion survey are not sufficient to make conclusions. However, the studies give first insights about the potential of the recommender.

It is worth noting that the current version of the recommender has been developed considering only academic programs and courses from the UOC, but it can be extended to cover other educational institutions. Interoperability structured exchange formats have been developed to easily add information from programs and courses of other higher education institutions.

Apart from improving the information extraction, it would also be interesting to improve the way information is represented to improve interoperability. The desirable scenario would be to have shared models that represent the necessary information of resumes and curricula. These models should be represented in a way that can be understood by programs: in an ontology framework for example. Our plan is to reuse the standardization proposals commented in the related research section to create an ontological framework. The proposed ontologies will act as a federated model that stores the information of the academic programs of different universities and the people resumes. Linked data technologies will be used since it is very difficult to find out a model that suits all domains and countries. So, even when we have different curricula or resume models (one for each country for example) their component classes and resources could be linked by using linked data techniques in order to act as a centralized database with all the information of the academic programs to consider.

From an experimental perspective, we are planning to use the collected comments from the survey to improve the system and to give better support to students. We are planning with the university employment service to enlarge the study to more students in different specialties. However, this will be performed with the improved version of the system.

From an institutional perspective, we are interested in including the generated information of the system within the institution analytical system. Adding the new data in the institutional data warehouse could help to improve the decision taking in the institution at management and teaching level. For example, they could be used to address labor needs in the design of new learning contents and therefore provide courses and programs more professional oriented.

6 Acknowledgment

This work was funded by the eLearn Center from the UOC, the SmartLEARN and SOM research groups, and the Spanish Government through the project: TIN2013-45303-P "ICT-FLAG: Enhancing ICT education through Formative assessment, Learning Analytics and Gamification".

7 References

- [1] Nankani, E., Simoff, S., Denize, S. & Young, L. (2009), Enterprise university as a digital ecosystem: Visual analysis of academic collaboration, in 'Digital Ecosystems and Technologies, 2009. DEST'09.3rd IEEE International Conference on', IEEE, pp. 727–732.
- [2] Taylor, E. (2012), Teaching data warehouses in collaboration with industry: Bridging the expectations gap, in 'Information Society (i-Society), 2012 International Conference on', IEEE, pp. 366–369.
- [3] Avdeenko, T. & Bakaev, M. (2014), Modeling information space for decision-making in the interaction of higher education system with regional labour market, in 'Actual Problems of Electronics Instrument Engineering (APEIE), 2014 12th International Conference on', IEEE, pp. 617–623.
- [4] Yankova, M., Saggion, H. & Cunningham, H. (2008), A Framework for Identity Resolution and Merging for Multi-source Information Extraction, a 'LREC', European Language Resources Association.
- [5] Bizer, C., Heese R., Mochol M., Oldakowski, R., Tolksdorf, R., Eckstein R. (2005). The Impact of Semantic Web Technologies on Job Recruitment Processes. International Conference Wirtschaftsinformatik (WI 2005), Bamberg, Germany. https://doi.org/10.1007/3-7908-1624-8_72
- [6] Schmidt, A., Kunzmann, C. (2006) Towards a Human Resource Development Ontology for Combining Competence Management and Technology-Enhanced Workplace Learning A Proceedings of OntoContent 2006, Springer, Lecture Notes in Computer Science (LNCS), vol. 4278, pp. 1078-1087.
- [7] Draganidis, F., Chamopoulou, P., Mentzas, G. (2006) An ontology based tool for competency management and learning paths. 6th International Conference on Knowledge Management (I-KNOW 06), Special track on Integrating Working and Learning.
- [8] Mansourvar, M. (2011) Development of a Job Web Portal to Capture Industry's Needs. Phd Dissertation. University of Malaya.
- [9] Bizer, C., Heath, T., & Berners-Lee, T. (2009). Linked data-the story so far. Semantic services, interoperability and web applications: emerging concepts, 205-227. <https://doi.org/10.4018/jswis.2009081901>
- [10] Vasiliev, V., Kozlov, F., Mouromtsev, D., Stafeev, S., & Parkhimovich, O. (2016). ECOLE: an ontology-based open online course platform. In Open Data for Education (pp. 41-66). Springer International Publishing. https://doi.org/10.1007/978-3-319-30493-9_3
- [11] Mikroyannidi, E., Liu, D., & Lee, R. (2016). Use of Semantic Web Technologies in the Architecture of the BBC Education Online Pages. In Open Data for Education (pp. 67-85). Springer International Publishing. https://doi.org/10.1007/978-3-319-30493-9_4
- [12] Zablith, F. (2015, May). Interconnecting and enriching higher education programs using linked data. In Proceedings of the 24th International Conference on World Wide Web (pp. 711-716). ACM. <https://doi.org/10.1145/2740908.2741740>
- [13] Taibi, D., Fulantelli, G., Dietze, S., & Fetahu, B. (2016). Educational Linked Data on the Web-Exploring and Analysing the Scope and Coverage. In Open Data for Education (pp. 16-37). Springer International Publishing. https://doi.org/10.1007/978-3-319-30493-9_2
- [14] Carreras, X., Chao, I., Padró, L., & Padró, M. (2004). FreeLing: An Open-Source Suite of Language Analyzers. Proceedings of the 4th International Conference on Language Resources and Evaluation (LREC'04).
- [15] Guitart, I., Conesa, J., Baneres, D., Moré, J., Duran, J., & Gañan, D. (2016). Extraction of Relevant Terms and Learning Outcomes from Online Courses. International Journal of

- Emerging Technologies in Learning (IJET), 11(10): 22-30 <https://doi.org/10.3991/ijet.v11i10.5928>
- [16] Feldman, R. & Sanger, J. (2006). The Text Mining Handbook: Advanced Approaches in Analyzing Unstructured Data. Cambridge University Press. <https://doi.org/10.1017/cbo9780511546914>
- [17] Miner, G., Delen, D., Elder, J., Fast, A., Hill, T. & Nisbet, R. (2012). Practical Text Mining and Statistical Analysis for Non-Structured Text Data Applications. Elsevier.
- [18] Moré, J., Conesa, J., Baneres, D., & Junyent, M. (2015). A Semi-Automated System for Recognizing Prior Knowledge. International Journal of Emerging Technologies in Learning (IJET), 10(7), pp. 23-30. <https://doi.org/10.3991/ijet.v10i7.4610>

8 Authors

David Baneres is with IT, Multimedia and Telecommunication Department at the Open University of Catalonia, Rambla del Poblenou 156, Barcelona 08018, Spain. His research interest involves e-learning innovative methodologies and intelligent tutoring systems. He is also a member of the SOM Research group at UOC (*som-research.uoc.edu*).

Jordi Conesa, is with IT, Multimedia and Telecommunication Department at the Open University of Catalonia, Rambla del Poblenou 156, Barcelona 08018, Spain. (e-mail: jconesac@uoc.edu). His research interest involves e-learning innovative methodologies, analytics in the educational context and intelligent tutoring systems. He is member of the SmartLearn research group at UOC (*smartlearn.uoc.edu*)

Article submitted 17 May 2017. Published as resubmitted by the authors 22 June 2017.