A Semi-Automated System for Recognizing Prior Knowledge

J. Moré, J. Conesa, D. Baneres and M. Junyent

Open University of Catalonia, Barcelona, Spain

Abstract—**Adaptive e-learning systems are able to automatically generate personalized learning paths from the students' profile. Generally, the student profile is updated with information about knowledge the student has acquired, courses the student has passed and previous work experience. Unfortunately, dealing with courses that students passed in other learning environments is very difficult, error prone and requires a lot of manual intervention. In addition, the recognition of external courses is a process that all institutions, on-site and online learning organization, must perform during the access of new students, since it can be greatly useful not only for personalization but also for recognizing the courses the students attended. In this paper, we propose an intelligent system that analyzes the academic record of students in textual format to identify what subjects the students studied in the past and therefore are potentially recognizable. In addition, the proposed system is able to enrich the information the institution has about the students' background, facilitating the identification of personalized learning paths.**

Index Terms— **Recognition, prior learning, support system, adaptive learning, background knowledge, context.**

I. INTRODUCTION

The aim of an adaptive e-learning system is to guide the learner to a comprehensive learning process based on the learner's knowledge [1]. Conceptually, adaptive learning can be performed in three different non-disjoint ways: 1) personalization of the learning processes: where the learning process of each student is personalized according to the student profile and behavior [2][3], 2) content personalization: where the learning objects (or the courses) offered to the student (and the activities the student should do) are personalized according to the student experience [4][5][6]; and 3) interface adaptation: where the learning management system interface is adapted according to every student's needs, preferences and history[7][8].

Adaptive systems should contain a great amount of information about students in order to select the content to be provided, along with the interface and the learning activities for each student. Such information is what we call the learning profile of the student. Adaptive e-learning systems tend to be self-contained, which means that the information that updates the student profile is automatically generated from the competences (or skills) the learner acquires in the virtual learning environment (VLE) [9][10]. Note that, many of these systems do not take into account the knowledge the learner acquired outside the institutional VLE. If so, they tend to use questionnaires [11][12] to find out about students' prior knowledge. Questionnaires may be useful to this task but they require student interaction and can be error prone. It would be useful for these systems to know what the student did outside the VLE automatically without having to ask students. In addition, questionnaires usually recognize prior knowledge of student by asking about content. However, we propose to recognize prior knowledge by finding out what content the students got in previous courses without using questionnaires. In that direction, the research question proposed in this work is: is it possible to create a knowledge-based system to automatically infer the knowledge students acquired from their previous courses? In order to do so, we propose to use the academic records of the previous courses the students attended.

This paper proposes an intelligent system to infer the main competences and knowledge the students have acquired from previous courses. The inference will be done by analyzing the academic records of students written in textual format. To do so Natural Language Processing (NLP from now on) techniques will be used. The contribution of the paper is double: 1) to speed up the recognition of prior learning by partially automating it, and 2) to update the students' knowledge profile by adding information about their previous learning.

The paper is structured as follows. Section II reviews the state of art related to prior learning recognition and it also describes the contribution proposed in this paper. The current official process of prior-learning recognition at our institution, which is a totally manual process, is presented in Section III. Section IV proposes a semi-automated system for prior learning recognition in any adaptive learning system; the ontology designed to model the system is introduced in Section V; and Section VI presents the prototype of an assistant for the semi-automated module. Finally, the experimental results are summarized in Section VII and the conclusions and future work are described in Section VIII.

II. BACKGROUND AND RELATED WORK

This section reviews the literature related to the evaluation of prior knowledge and it presents the aim of the paper. Several works have been published related to how to evaluate prior learning and what items must be assessed. The challenge is to detect all the learning background of a student based on the evidences that could be collected. The authors in [13] identified six types of assessment methods: multiple-choice tests, open questions/completion tests, association tests, recognition tests, free recall and self-assessment. However, they concluded that test-based assessment was the best option compared to self-assessment or free recall because they mostly depend on the student's verbal abilities to describe

his skills. The authors in [14] argued that the content of a test is more important than the format since different kinds of knowledge are detected depending on the performed questions. Based on these conclusions, several experiments in different areas were published, such as, in mathematics [15], in physical education [16] or even for excluding learners who do not possess the necessary skills to perform a course [17]. The results reported statistics of the quality of the prior knowledge extracted.

However, questionnaire-based evaluation is not the only technique that could be applied to collect prior learning. A student usually has academic reports and transcripts of all the education received. These reports are official documents that prove the knowledge acquired in educational institutions. Therefore, this documentation can be processed to extract all the prior learning.

In the university context, the process of evaluating the students' prior knowledge is regulated and defined as recognition of prior learning. This evaluation is an important process that any on-site or online university should perform when a new student is registered. The reason is to detect the subjects the student does not need to enroll according to the proved competences acquired in other official institutions.

There are also initiatives in several countries (Canada, United Kingdom, Australia, New Zealand) to recognize informal knowledge acquired by professional experience and employment. Different denominations such as Recognition of prior learning (RPL), prior learning assessment (PLA), or prior learning assessment and recognition (PLAR) are used depending on the country.

Such processes are hard, error-prone and usually performed manually, since all the documentation delivered by the student is in text format scanned as an image, which prevents an easy integration with VLE. It requires the participation of several technicians and lecturers of the university. The main drawbacks are the manual process, the uncertainty of the required time (required time is affected by the number of students who request the recognition process and the amount of the documentation delivered) and the amount of time required to answer students about the proposed recognitions. The procedure may also incur in recognition errors due to the large quantity of documentation to be analyzed.

The created system can be seen as a support module that can be added to any adaptive e-learning system and allows the semi-automatic identification of the prior knowledge the student has acquired. This knowledge can be used to recognize prior learning or to personalize learning paths for each student.

This paper focuses on a support system that proposes potential recognitions for each subject by using simple NLP techniques. The proposed system may improve the quality of the process by reducing dramatically the time the lecturers need to analyze the documentation. Such reduction is attained by giving advices that help to perform the task, and by finding out the most similar subjects to the ones in the academic records (certifications, etc.) provided by students. This paper presents the first step of a fully automatic system to update the learner profile from external evidences in an adaptive e-learning system.
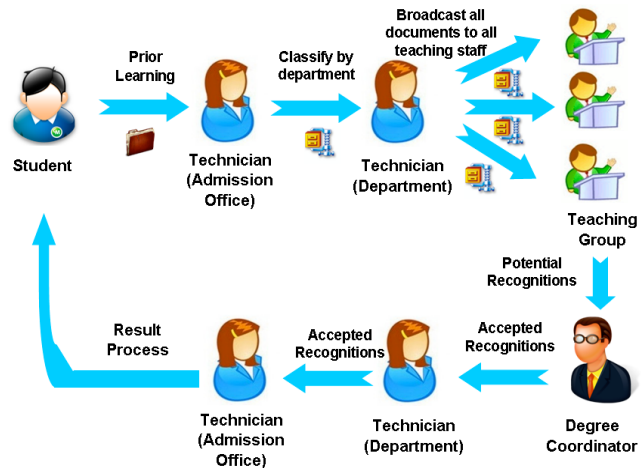

Figure 1: Previous manual recognition process

As proposed in [18], natural language processing (NLP) techniques can be used in order to semi-automate processes related to documentation analysis and, therefore, make them more sustainable. Some work was previously presented NLP techniques for processing documentation. Authors in [19] proposed NLP techniques to detect duplicate patterns on a documentary repository with descriptions of software code defects. The objective of the tool was to remove redundancy in the repository. Authors in [20] presented a tool that automatically catalogues educational digital resources according to their contents and the educational standards of the educational system in their country.

III. PROCESS OF RECOGNITION

This section describes the manual process to recognize prior learning in our institution. The objective of the section is to present a valid existing procedure to evaluate the previous knowledge of students. As we will discuss later, our suggestion tries to improve this process by automating the prior knowledge recognition.

The process involves several technicians and teaching staff (See Figure. 1). The student provides all the teaching learning plans of the completed subjects. Here, we define a teaching learning plan (or teaching plan) for a subject as an official document generated by the external institution where the course info, the competences to be acquired, the syllabus and the course schedule are described. All the documentation is digitalized (one image file for each subject) and archived in a compressed file (one compressed file for each student) in the admission office.

The digitalization process reduces significantly the paperwork. However, the information is stored in image format where the name of each file is the only external evidence of the subject described inside the document. Currently, there is no accepted procedure to receive digitally the documentation since all the documents must be certified (official stamp) by the external institution.

Then, these documents are sent to the respective department depending on the area of knowledge of the degree the student has been enrolled. After that, the technician of the department sends all the documentation to all the teaching staff related to the degree. The teaching staff is divided in groups of lecturers depending on the specialty of their subjects.
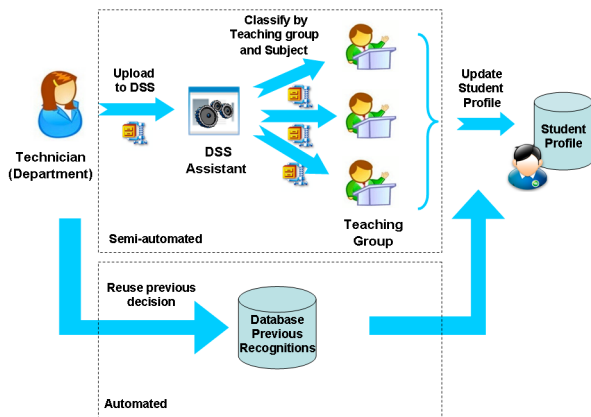
Figure 2: Semi-automated recognition system

The teaching staff has to process all the documents since there are no enough external evidences of the potential recognition (only the name of the image files that correspond to the name of the subject or its abbreviation). This handicap produces many errors in the recognition process and high workload because all the documents must be analyzed in order to evaluate whether a subject should be recognized. Afterwards, the potential recognitions are validated by the degree coordinator and the accepted documents are sent to the technician staff to update the academic record of the student. All this procedure requires approximately 3 months that could even be longer in case of a student allegation.

In this process, many errors may arise due to a large number of involved staff and documentation to be analyzed and the time constraints. This process has been proved as a successful process over time, however we believe that using NLP techniques to automate some parts of the process can make it more agile and efficient.

In the following lines the criteria used to recognize previous subjects are described.

A. Criteria of prior learning recognition

In Spain, the subjects of the degrees based on the European Higher Education Area (EHEA) are designed according to the competences the student should learn. Therefore, the recognition should be done according to the acquisition of these competences.

When the recognition is performed for another degree based on the EHEA, the evaluation is simpler. If two subjects have the same competences, they should be potentially recognized regardless their contents. Moreover, there is a special case of degree recognition for the same branch of knowledge. In this case, the subjects defined as basic in the first academic year are directly recognized independently of their contents.

However, the criteria change when the recognition is compared with degrees based on the previous Law of University Reform (LRU) [21] due to the impedance mismatch between them: the EHEA subjects are described by means of competences and LRU subjects are described by means of objectives.

IV. A SIMPLE KNOWLEDGE-BASED SYSTEM TO SEMI-AUTOMATE THE RECOGNITION OF PRIOR KNOWLEDGE

The manual process described in the previous section requires a high workload for technicians and teaching staff, which hinders a simple integration with any VLE. Our proposal adapts the official recognition process to evaluate the students' knowledge in an adaptive learning system. The criteria described in the previous section should be carefully modified in order to accept only recognitions based on the contents and workloads of the subjects. This modification will successfully improve the quality of the stored profile. In order to automate the recognition system, we propose the system illustrated in Figure 2.

When preprocessing a request of recognition, the technicians of the department can reuse all the previous decisions performed for the degrees of the same external institution in contrast with a degree of our university. That means that if a subject of a given university was recognized in the past, then the technician can automatically recognize that subject on new requests. Using previous knowledge, the number of manual recognitions can be drastically decreased.

For any not-yet-analyzed degree, the process can be performed by using a decision support system that assists technicians and academic staff in the recognition process. That assistant would semi-automate and speed up the recognition of external degrees. In particular, the assistant should help the department technicians to identify the potential subjects to be recognized, the specialty of the subjects and, therefore, the teaching group of the specialty the documentation should be sent to according to the detected content.

The proposed system should analyze the documents presented by a student and extract the most relevant content elements. Afterwards, these elements are compared with the curricula of the degree the student has been enrolled to. As a result, similarity scores are generated; estimating how similar is the provided subject against all the subjects of the enrolled degree. Then, all this information is analyzed, the potential subjects to be recognized are filtered, and they are sent to the teaching staff. From this point, the process becomes manual as the previous recognitions process, since the recognition decision is performed manually.

Therefore, the main benefits of the proposed approach are: 1) the ontology that stores information about past recognitions and context information, and 2) the knowledge-based assistant that uses such information to identify the potential subjects to recognize. The next two sections focus in both benefits, Section V presents the ontology used to represent previous recognitions and Section VI describes on more detail how the assistant of the semi-automated module works.

V. ONTOLOGY MODEL

This section describes the ontology model proposed for the semi-automated system. Note that the design has been performed from scratch based on the necessary knowledge to be stored.

Several ontology-based approaches for modeling the student profile have been proposed. Independent relevant designs [22][23][24] can be found in the literature. Related to the learner information, two well-developed standards were specified: IEEE PAPI [25] and IMS LIP [26]. PAPI stresses on the student performance information, whereas IMS LIP focuses on the classical notion of curriculum vitae. Following these standards, other ontology models have also been proposed [27][28].
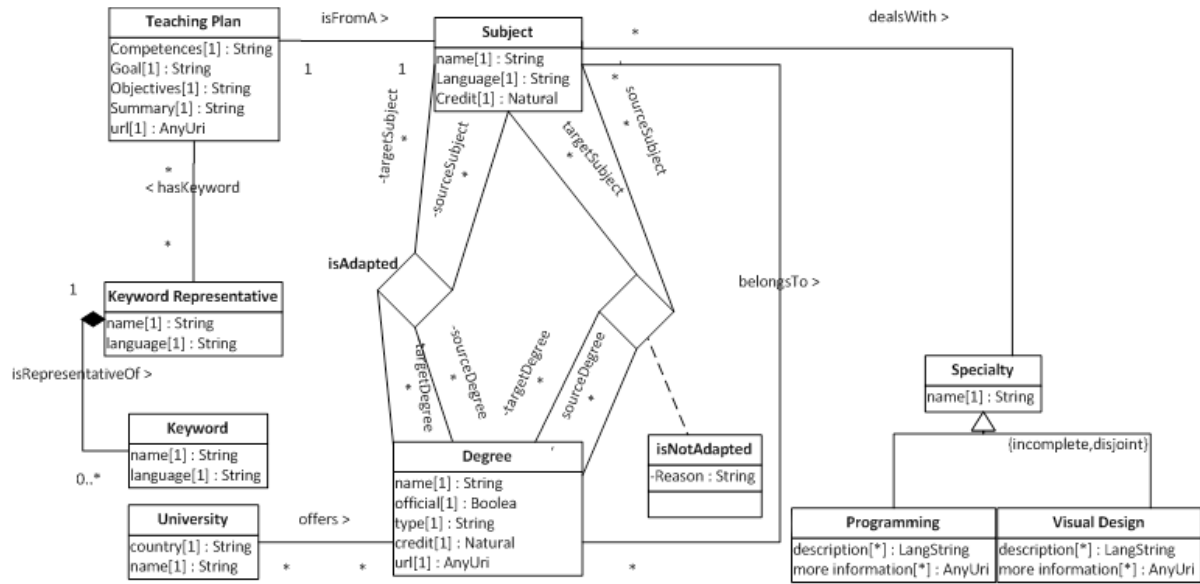
Figure 3: Ontology model of the semi-automated system

Our system deals with information related to the recognition between degrees. Therefore, we decided to create a new ontology capable of being integrated in the future in any of the previous models. Thus, for each recognized subject, the competences the student has acquired can be easily obtained. The reader may wonder why using an ontology. The reason lies in using the past knowledge, despite the ontology is actually used as a database, to improve the quality of the assistant recognitions in short term.

The proposed ontology contains the information required to automatically recognize subjects (the subjects that have been recognized in the past) and to improve the identification of the potential subjects to recognize (the proposed but non-recognized subjects, the teaching plans of each processed subject and their automatically extracted keywords, the information about the different specialties of the university and the teaching groups that coordinate each specialty).

A UML representation of the ontology can be seen in Figure 3. The main class of the ontology is *subject*. It is important to identify the credits (that denotes the minimum workload required by students) of each subject since it is necessary for the recognition process. Otherwise, it would be impossible to detect that a subject that requires 150 hours to complete cannot be recognized by a subject that only requires 60 hours. Each subject has a teaching plan that describes all the aspects of the subject by using natural language. Even though a teaching plan can contain multiple information, the proposed system only extracts and stores information related to the competences achieved through the subject, the goals the subject has, a summary and the URL where it can be downloaded. Currently the competences information is stored as plain text to help the NLP techniques to find coincidences within subjects. However, this information should meet in the future specification standards as IEEE RCD [29] and IMS RDCEO [30] in order to be able to link with existing models and, even, recognize prior learning from systems where the learning outcomes of students are stored using compatible models.

Ontology also stores information about the extracted keywords of each subject. Note that in some cases, different teaching plans can use different keywords to denote the same concept. Such information is also stored in the ontology by using the *keyword representative* class. This class allows to group keywords that share the same meaning (i.e. "business intelligence", "BI", "inteligencia de negocio", etc.) and to define the keyword of the group that should be used as a representative (i.e. "BI").

The same subject may belong to one or more degrees (see relationship *belongsTo*). For example, a subject titled *Foundations of Calculus* may belong to a degree of *Computational Mathematics* and also to a degree of *Information Systems*. The quaternary relationship types *isAdapted* and *isNotAdapted* determine the accepted and rejected recognitions, respectively. In case of rejection, the ontology stores the reason of rejection. Note that, the relationship types need to be quaternary, since the recognition should be taken into account in the context of a given degree. A subject that belongs to several degrees may be recognized in one degree but not in the others. Following the previous example, the subject *Foundations of Calculus* may be recognized for another subject in the context of the degree of *Information Systems* but not in the context of the *Computational Mathematics* degree. The reason is that *calculus foundations* is more relevant in the context of *Computational Mathematics* and therefore the criterion in recognizing the subject in such context is more demanding than in *Information Systems*.

Finally, it is also relevant that the ontology represents the different specialties (area of knowledge) dealt by the target university and the teaching group responsible of each specialty. Such information would facilitate the identification of the domain experts that are responsible of validating the recognitions for each subject.

VI. SUPPORT SYSTEM

In this section, we describe all the technical aspects related to the design of the assistant. The proposed system is composed by 5 modules as it is illustrated in Figure 4: (A) Automatic analysis of curricula from our university: periodically the subjects of our university will be analyzed
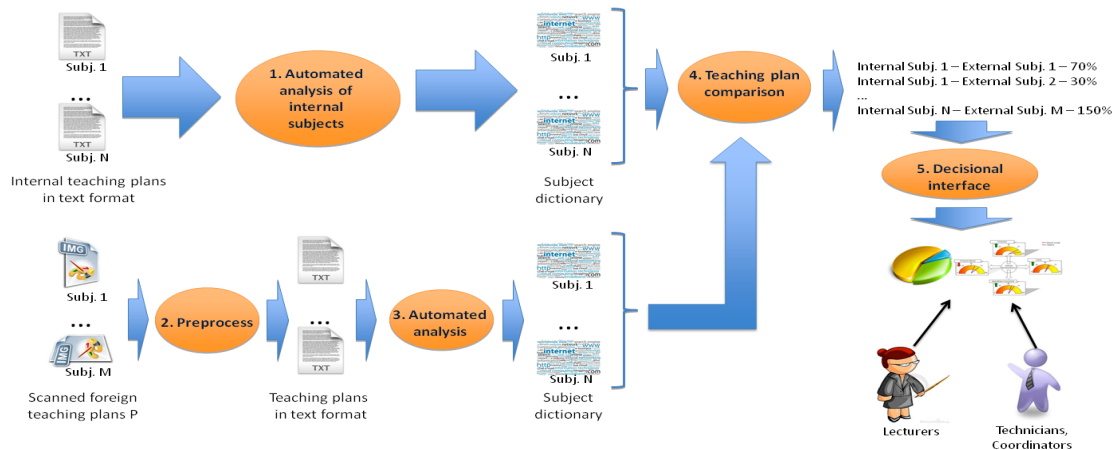
Figure 4: Diagram of the teaching plan recognition assistant

in order to create a dictionary of relevant keywords for each subject. (B) Normalization of the presented documents and conversion to text: the documents presented by the students will be processed to be normalized. Such process will deal with the change of resolution of the documents, their change of orientation and their conversion to text using an Optical Character Recognition (OCR). (C) Analysis of the provided documents: a dictionary of relevant keywords will be created for each of the provided documents. (D) Comparison of internal and external curricula: the subjects provided by students will be compared with the subjects of the enrolled degree by means of their dictionaries. (E) Decisional Interface: the obtained information will be presented in a way to support the teaching staff in the recognition, by filtering the relevant information for each teacher and providing the information using dashboards.

The next subsections describe in more detail each module of the system.

A. Analysis of own degrees

The relevant content of each own degrees must be extracted, in order to compare the content of the degrees of our university with the content of the degrees in the teaching plans from external universities. Each semester, an automatic process analyses the own university teaching plans by using NLP methods. The goal of the analysis is to create one dictionary for each subject with the main keywords of the subject, which we call *dictionary of subject domain terms* (DSDT).

The main keywords are extracted by first calculating the n-grams of the teaching plan content. An n-gram is a sequence of subsequent words in a text. For instance, in the text *Differential equations in those material environments*, sequences such as *differential equations*, *equations in*, *in those material*, *those material environments*, and *material environments* are n-grams.

In order to obtain the keywords, the n-grams are filtered according to the following criteria: (1) Combinations of words with certain parts of speech are not allowed; and (2) n-grams which are not significantly relevant in the document. The first criterion filters out n-grams with *stop words* (prepositions at the beginning and the end, adverbs, pronouns, conjunctions and determiners). Thus, some of the n-grams filtered out in the above text are: *equations in*, *in those material*, *those material environments*. The second criterion calculates the term frequency–inverse document frequency (TF-IDF) of the n-grams that remain

and are equal to Wikipedia titles. The TF-IDF calculation is performed by contrasting the appearance of the n-gram in the teaching plan content with its appearance in a collection of representative documents of the use of the language in different domains (news about sports, politics, culture, etc). N-grams with a large TF-IDF score are terms which are not of general use and they are more specific to the teaching plan content. Besides, if the n-gram matches a Wikipedia title, the n-gram is more likely to be a relevant keyword with some academic meaning. As an example, the dictionary of the subject *Numerical analysis* (from the degree in Computer Engineering) will register n-grams such as *differential equations*, or *Gauss theorem*, in contrast to n-grams such as *equations*, *evaluation*, or *knowledge consolidation*.

The system currently processes the whole teaching plan to get all the relevant terms, not only the subject competences. Note that the assistant does not infer the workload or the schedule of the subject. Currently, these evidences for potential recognition should are still manually identified.

B. Preprocessing of external teaching plans

As aforesaid, the external teaching plans are scanned and delivered as image files (stored in PDF format), one for each subject. These image files must be processed in order to extract its text. The OCR program used to do so is the TesseractOCR (http://code.google.com/tesseractocr)

In order to get minimally satisfactory results in a reasonable period of time, the documents must be vertically oriented and the image resolution cannot be excessively high. In order to guarantee so, the following automatic processes are activated: (1) Document orientation unification: the documents can be delivered in any orientation (horizontal, vertical, two-pages per sheet with vertical orientation, etc.) The goal is to ensure that the documents to be processed by the OCR have the same orientation (vertical). A set of algorithms are run to identify the orientation of each document and change when necessary. (2) Change PDF document resolution: the content in the PDF files are in image format. After several tests, we noticed that the OCR results were significantly worse when the resolution was below 150 dpi (dots per inch). When the resolution was set above 200 dpi, the process was extremely slow with no significant better results. So this process applies the 200 dpi resolution in order to guarantee an acceptable trade-off between time and quality results. (3) OCR application: a

Figure 5: Decision interface for a subject filtered by specialty

set of text format files are obtained, one for each subject. These files contain the teaching plans previously scanned in plain-text format. (4) Data cleansing: possible misinterpretations are corrected by using dictionaries and edit distance algorithms that calculate the shortest distance between an incorrect and a correct word. Notice that, the result of these processes, i.e. the keywords in the DSDT, highly depends on the quality of documents and the OCR program.

C. Analysis of external teaching plans

Once the plans delivered by students are in plain-text format, the subject keywords are identified. The identification process is the same as the one that identifies the keywords of the degrees of our university (see Subsection *VI.A*). The outcome is a DSDT dictionary for each subject. This dictionary contains the most representative and defining terms of the subject.

D. Comparison of teaching plans

The following stage consists in comparing the external subjects presented by students with the subjects of our university. This comparison is performed by calculating the term coverage of both the internal and the external subject DSDT dictionaries. For each term in the external DSDT, the process registers the internal DSDTs that contain this term. The subject of our university whose DSDT has the largest number of matched terms is the one considered the most alike.

The result is a sorted list according to the number of matched terms in their DSDT dictionaries that indicates the likeness between the subject presented by the student and each subject of our university. For example, the external subject Optical Communications may have terms that highly match those of own subjects like mobile communications, Mobile and wireless networks and Telematic systems.

E. Decision Interface

The final stage analyses the information obtained and displays this information in a way that facilitates users (teachers and program coordinators) to take decisions. It is important to establish the content and the format to be displayed according to the user type.

The system allows the user to filter the information according to the consulted subject (See Figure 5). For instance, if the user wants to know whether the external subjects are similar to a specific internal subject, the subjects of other degrees are deleted and those internal subjects which are thematically more similar are focused. The system finally promotes the subjects with more terms in common with the external teaching plan. The number of promoted subjects can be configured (5 subjects by default in our experiments). Note that, for each subject the original teaching plans and common terms can be also

reviewed for the final decision. Let us imagine that the external subject Physics I is similar to the internal subjects Mathematics II, Physical Principles of Engineering, Statistics, Game theory applied to political sciences and Mathematics for Engineering, among others. If the user wants to check whether the external subject can be validated in the Computing Engineering degree, the alien subjects, such as Game theory applied to political science will be disregarded. From the remaining list, the subject Physical principles of Engineering will be promoted because this subject and Physics I share terms like Gauss Theorem and Maxwell.

The assistant also performs recommendations at a higher granularity level: teaching groups. Teaching groups, as said in Section III, are groups that manage the subjects of a given specialty area. This difference in granularity level can be very convenient because recommendations focused on subjects may be error prone. Furthermore, the distribution of contents through different subjects change in different universities and the differences between similar subjects may be diffuse when applying NLP techniques. Recommendations based on teaching groups increase the precision of the assistant.

As regards the presentation format, the system currently displays the following output, according to the final user: (1) Recommendation of the subjects more likely to be validated (a higher number of common terms), and (2) Reports that support and guarantee the process.

The tool has been integrated as an experimental module in the management recognition system in our university. The support system currently works manually. When the technician of the department receives the new documentation, the documents are uploaded to the system and processed in batch mode. The output is in HTML mode with the information described in this section (See Figure 5). As described in Section III.A, the result of the recognition is processed manually as in the previous process. Note that, in future versions the system should be automated and fully integrated with the recognition procedure.

VII. EXPERIMENT

This section presents experimental results performed by the designed prototype of the support system. The assistant has been tested in a pilot in order to evaluate the level of correctness of the results. The manual process and the assistant process were performed at the same time, and the results were compared. The objective is to test the recognized subjects and the identification of the teaching group who has to evaluate the recognition.

The experiment involved the automatic analysis of 52 external subjects to be recognized from the degree of *Telecommunication Engineering* at our university (Universitat Oberta de Catalunya). The evaluation of the reliability of the results has been carried out by comparing the manual recognition by the teaching staff with the proposed results from the assistant.

The manual process recognized the 8 subjects presented in TABLE I. Out of the 44 rejected subjects, 38 were discarded because of a non-matching curriculum in the degree and 6 because of a non-matching subject in the degree. TABLE I shows that the assistant finds the correct teaching group in 100% of the 8 recognized subjects. The assistant proposes the correct group alone in five of the

TABLE I.
RELATION BETWEEN RECOGNIZED SUBJECTS AND THE RECOMMENDATION PERFORMED BY THE ASSISTANT. THE TABLE SUMMARIZES THE EXTERNAL SUBJECTS, THE RECOGNIZED SUBJECTS, THE TOP FIVE PROPOSED POTENTIAL SUBJECTS AND THE TEACHING GROUP PROPOSED BY THE ASSISTANT. THE SUBJECTS AND SPECIALITIES THAT MATCH THE MANUAL RECOGNITION PERFORMED BY THE TEACHING STAFF ARE UNDERLINED.

External Subject	Recognized Subject	Potential Subjects to be Recognized	Proposed Speciality
Operating systems	Operating systems	Operating systems	Computer architecture
Network, analysis and synthesis	Technology Fundamentals	Circuit theory, Mathematics II Digital electrical systems, Technology fundamentals II, Mathematics fundamentals	Applied mathematics, statistics and operational research, Electrical technology, signal theory and communications
Laboratory of data communication	Network structures	Operating systems, Advanced administration of GNU/Linux operating systems, Internet protocols and applications, Networks, Network structures	Computer architecture Computer networks
Algebra	Mathematics I	Mathematics I, Mathematics fundamentals, Mathematics in Computer Science, Mathematics II, Algebra	Applied mathematics, statistics and operational research
Programming	Programming fundamentals	Programming fundamentals, Digital electrical systems, Operating systems, Advanced administration of GNU/Linux operating systems, Networks and operating systems administration	Electrical technology, signal theory and communications, Software Engineering and programming
Economy	Introduction to the economy	Introduction to the economy	Business Organization.
Advanced mathematics	Mathematics II	Mathematics II, Mathematics I, Mathematical analysis, Mathematics fundamentals, Signals and systems II	Applied mathematics, statistics and operational research
Physics I	Physics fundamentals in Computer Science	Mathematics I, Physics fundamentals in Computer Science, Mathematics fundamentals, Mathematics in Computer Science, Networks and Services	Applied mathematics, statistics and operational research

subjects. If the recognized subject is compared with the potential list of recognitions, the ratio of success decreases to 87% (it fails in *Network, analysis and synthesis*). Among the other subjects, the correct one is in the list of the 5 potential recognitions. Note that this result will be frequent in this assistant due to the ambiguity in some subject syllabus.

Related to the non-matching contents, the assistant produced some false positives in 34% of the cases. In the subjects with no correspondence in the degree, the assistant proposed a teaching group and a list of potential recognitions. In the subjects with no coincident content, the assistant returned an incorrect teaching group for 9 subjects. The reason of these incorrect results could be related to subjects with few and common terms. For instance, the term *statistics* is a common term that appears in the subject *Statistics for the linguistic investigation*. Then, the assistant proposed the subject *Statistics I*. Other possible errors could appear during the OCR recognition or due to the low quality of the some documents.

Globally, the 56% of the subjects were properly selected. Although, the precision of the assistant is not accurate, it reduces considerably the total workload of the lecturers. In this case, the assistant properly selected the correct subjects for 12 out of the 15 teaching groups of the degree. That is a great improvement since, in most cases, the staff of 14 teaching groups will not have to analyze irrelevant documents.

VIII. CONCLUSIONS AND FUTURE WORK

The paper proposes a semi-automatic system to evaluate previous student's knowledge. The system could be used to evaluate prior knowledge in course recognition or to create the initial student's profile in an adaptive learning system.

The system includes an assistant to help the teaching staff to speed up the evaluation of the documentation provided by the student and decrease the errors appeared in the previous manual process. The experimental results obtained by the assistant are promising. In the level of subject recognition the precision is around 50%, but when dealing at the teaching group granularity the system performed with a precision of 100%.

As future work, we propose to refine the assistant to improve the quality of the results reducing false positives and to perform a larger experiment on degrees of different areas of knowledge. Moreover, the experiments have been performed only on Spanish and Catalan degrees where the language is similar in terms of NLP techniques. The modelled ontology already supports the internationalization of the process. However, the NLP techniques have not been yet implemented. We propose to add a language detector module on the analysis of the teaching plan in order to be able to apply the recommender worldwide. Additionally, the system currently only supports recognition of previous knowledge from learning courses. However, it should be interesting to add a recognition module based on professional competences [31]. The system should use standard models such as Professional Learning Ontology and Competencies (PLOC) [32] and General Competency Ontology (GCO) [33] to specify the competencies.

Finally, we will analyse how this system could be integrated in any adaptive learning system. In terms of the ontology model, the information related to the competences acquired can be easily extracted from the recognized subjects. Therefore, this information can be added to any student profile model.

REFERENCES

[1] S. Chen and G. Magoulas, "Adaptable and Adaptive Hypermedia Systems" Hershey, PA: IRM Press, 2005. http://dx.doi.org/10.4018/978-1-59140-567-2

[2] L. Huey-Ing and Y. Min-Num, "QoL guaranteed adaptation and personalization in E-learning systems," Education, IEEE Transactions on, vol.48, no.4, 2005, pp.676-687. http://dx.doi.org/10.1109/TE.2005.858398

[3] Sfenrianto and H. Suhartanto and Z.A. Hasibuan, "A dynamic personalization in e-learning process based on triple-factor architecture," Computing Technology and Information Management (ICCM),International Conference on, 2012, pp.69-75

[4] A.-E. Guerrero-Roldán, I. García-Torà, J, Prieto-Blázquez, and J. Minguillón, "Using an IMS-LD based questionnaire to create adaptive learning paths," *Frontiers in Education Conference (FIE), 2010 IEEE* , 2010, pp.F1J-1,F1J-6

[5] N. Capuano and M. Gaeta and S. Salerno and G.R. Mangione, "An Ontology-Based Approach for Context-Aware E-learning," Intelligent Networking and Collaborative Systems (INCoS), 2011 Third International Conference on, 2011, pp.789-794

[6] N. Capuano and S. Salerno and G.R. Mangione and A. Pierri, "Management of Latent Learning Needs in Adaptive e-Learning Systems," Complex, Intelligent, and Software Intensive Systems (CISIS), Seventh International Conference o*n*, 2013, pp.361-366.

[7] J. Liu and C. K. Wong and K. K. Hui, "An adaptive user interface based on personalized learning," Intelligent Systems, IEEE, vol.18, no.2, 2003, pp.52-57. http://dx.doi.org/10.1109/MIS.2003.1193657

[8] Y. Fan and D. Teng and H. Yang and C. Ma and G, Dai and H. Wang. "An Adaptive User Interface Model Based on Experience Awareness", Chinese Journal of Computers, vol.34, no.11, 2011, pp.2011-2211. http://dx.doi.org/10.3724/SP.J.1016.2011.02211

[9] M. Yaghmaie, and A. Bahreininejad, "A context-aware adaptive learning system using agents", Expert Systems with Applications, vol. 38, no. 4, 2011, pp. 3280-3286. http://dx.doi.org/10.1016/j.eswa.2010.08.113

[10] M. Siadaty, and F. Taghiyareh, "PALS2: Pedagogically Adaptive Learning System based on Learning Styles," Advanced Learning Technologies, 2007. ICALT. Seventh IEEE International Conference on, 2007, pp.616-618.

[11] V. Esichaikul, S. Lamnoi, C. Bechter, "Student Modelling in Adaptive E-Learning Systems", Knowledge Management & E-Learning: An International Journal,vol. 3, no. 3, 2011, pp 342-355.

[12] J. C.R. Tseng, H.-C. Chu, G.-J. Hwang, C.-C. Tsai, "Development of an adaptive learning system with two sources of personalization information",Computers&Education,vol.51,no.2,2008,pp 776-786. http://dx.doi.org/10.1016/j.compedu.2007.08.002

[13] F.J.R.C. Dochy, M. Segers and M.M. Buehl, "The relation between assessment practices and outcomes of studies: The case of research on prior knowledge". Review of Educational Research, vol. 69, no. 2, 1999, pp. 145–186. http://dx.doi.org/10.3102/00346543069002145

[14] A. M. Shapiro. "How including prior knowledge as a subject variable may change the outcomes of learning research". American Educational Research Journal, vol.41, no.1, 2004, pp. 159–189. http://dx.doi.org/10.3102/00028312041001159

[15] T. Hailikari and A. Nevgi and S. Lindblom-Ylänne, "Exploring alternative ways of assessing prior knowledge, its components and their relation to student achievement: a mathematics based case study". Studies in Educational Evaluation, vol.33, no.3–4, 2007, pp.320-337. http://dx.doi.org/10.1016/j.stueduc.2007.07.007

[16] S., Bo and A. Chen. "An examination of learning profiles in physical education." Journal of Teaching in Physical Education vol.26, no.2, 2007, pp.145-160.

[17] P. McGee. "Distance learning supports: The intentionality of pre-assessment surveys". Proc. of World Conference on E-Learning in Corporate, Government, Healthcare, and Higher Education, 2002, pp. 1887-1890.

[18] P, Jackson and I. Moulinier, "Natural Language Processing for Online Applications: Text Retrieval, Extraction, and Categorization". 2nd Edition, John Benjamins Pub. 2007.

[19] Q. Wu, and Q. Wang, "Natural Language Processing Based Detection of Duplicate Defect Patterns". Computer Software and Applications Conference Workshops (COMPSACW), IEEE 34th Annual, 2010. pp. 220–225. http://dx.doi.org/10.1109/compsacw.2010.45

[20] H. Devaul, A. R. Diekema, and J. Ostwald, "Computer-assisted assignment of educational standards using natural language processing". Journal of the American Society for Information Science and Technology, vol. 62, no. 2, 2011, pp. 395–405. http://dx.doi.org/10.1002/asi.21437

[21] Spanish Government, "Organic law of University Reform (LRU) 11/1983, august 25th, 1983". BOE 209 September 1st 1983.

[22] J. Kay, "Ontologies for reusable and scrutable student model," position paper, in AIED99 Workshop on Ontologies for Intelligent Educational Systems, 1999, pp. 72-77.

[23] L.S. Muñoz and J. Palazzo and M. Oliveira, "Applying Semantic Web Technologies to Achieve Personalization and Reuse". In Proc. of the SWEL Workshop at Adaptive Hypermedia, 2004, pp.348–353.

[24] J. Clemente and J. Ramirez and A. de Antonio, "A proposal for student modeling based on ontologies and diagnosis rules". Expert Systems with Applications Journal, 2011, pp.8066–8078.

[25] Draft Standard for Learning Technology - Public and Private Information (PAPI) for Learners (PAPI Learner) - Core Features, IEEE P1484.2.1/D8, 2001. http://metadata-standards.org/Document-library/Meeting-reports/SC32WG2/2002-05-Seoul/WG2-SEL-042_SC36N0175_papi_learner_core_features.pdf.

[26] IMS Learner Information Packaging Information Model Specification, version 1.0, 2001. Available: http://www.imsglobal.org/profiles/lipinfo01.html.

[27] D. Paneva, "Use of Ontology-based Student Model in Semantic-oriented Access to the Knowledge in Digital Libraries", In Proc. of HUBUSKA Fourth Open Workshop Semantic Web and Knowledge Technologies Applications", 2006, pp. 31-41.

[28] K. Rezgui and H. Mhiri and K. Ghédira, "An Ontology-based Profile for Learner Representation in Learning Networks". International Journal of Emerging Technologies in Learning (iJET)**,** vol. 9, no. 3, 2014, pp. 16-25. http://dx.doi.org/10.3991/ijet.v9i3.3305

[29] IEEE RCD 2007. IEEE LTSC Reusable Competency Definitions. http://www.doleta.gov/usworkforce/pdf/2007-ieeecomp.pdf

[30] A. Cooper and C. Ostyn, C. "IMS Reusable Definition of Competency or Educational Objective", http://www.imsglobal.org/competencies/rdceov1p0/imsrdceo_infov1p0.html

[31] M. A. Sicilia, "Ontology-based competency management: Infrastructures for the knowledge intensive learning organization". Lytras, M.D., Naeve, A. (Eds.) Intelligent Learning Infrastructures in Knowledge Intensive Organizations: A Semantic Web Perspective, Idea Group, 2005, pp. 302-324.

[32] A. Schmidt and C. Kunzmann, "Towards a Human Resource Development Ontology for Combining Competence Management and Technology-Enhanced Workplace Learning" In Proc. Intern. Conf. on On the Move to Meaningful Internet Systems. 2006, pp. 1078-1087 http://dx.doi.org/10.1007/11915072_10

[33] J. M. Dodero, S. Sanchez-Alonso and D. Frosch-Wilke, D. Generative "Instructional Engineering of Competence Development Programmes". Journal of Universal Computer Science, vol. 13. no. 9, 2007, pp. 1213-1233

AUTHORS

J. Moré is with Learning Technologies at the Open University of Catalonia, Rambla del Poblenou 156, Barcelona 08018, Spain. (e-mail: jmore@uoc.edu).

J. Conesa, is with IT, Multimedia and Telecommunications Department at the Open University of Catalonia, Rambla del Poblenou 156, Barcelona 08018, Spain. (e-mail: jconesac@uoc.edu).

D. Baneres, is with IT, Multimedia and Telecommunications Department at the Open University of Catalonia, Rambla del Poblenou 156, Barcelona 08018, Spain. (e-mail: dbaneres@uoc.edu).

M. Junyent, is with IT, Multimedia and Telecommunications Department at the Open University of Catalonia, Rambla del Poblenou 156, Barcelona 08018, Spain. (e-mail: mjunyent@uoc.edu).