A Study on the Possibility of Automatically Estimating the Confidence Value of Students' Knowledge in Generated Conceptual Models

Diana Pérez-Marín, Enrique Alfonseca, Pilar Rodríguez and Ismael Pascual-Nieto
Department of Computer Science, Universidad Autónoma de Madrid, Spain
Email: {diana.perez, enrique.alfonseca, pilar.rodriguez, ismael.pascual}@uam.es

***Abstract*— We propose a new metric to automatically evaluate the confidence that a student knows a certain concept included in his or her conceptual model. The conceptual model is defined as a simplified representation of the concepts and relationships among them that a student keeps in his or her mind about an area of knowledge. Each area of knowledge comprises several topics and each topic several concepts. Each concept can be identified by a term that the students should use. A concept can belong to one topic or to several topics. Terms are automatically extracted from the answers provided to an automatic and adaptive free-text scoring system using Machine Learning techniques. In fact, the conceptual model is fully generated from the answers provided by the students to this system. In the paper, the automatic procedure that makes it possible is reviewed in detail. Finally, concept maps are used to graphically display the conceptual model to teachers and students. In this way, they can instantly see which concepts have already been assimilated and which ones should still be reviewed.**

***Index Terms*— metrics of concept assimilation; generation of conceptual models; free-text scoring; blended learning; e-assessment; e-learning**

I. INTRODUCTION

It is commonly agreed that knowledge, in part, is expressed in concepts. People use concepts in their daily lives. Many definitions have been provided for concept. The one chosen for this work is that a concept is "*a perceived regularity in events or objects, or records of events or objects, designated by a label*" [1]. From this definition, two important features of a concept can be identified: its archetype nature and its necessity of being denominated. In fact, it is important to notice that all concepts need to be assigned a label, something that identifies them since without this label the concepts are inaccessible.

Words are labels that map these concepts onto our knowledge structure. However, not all words serve to convey concepts since some of them express actions or links. In this work, a special relevance will be given to the words that express concepts (usually nouns or group of nouns) and their label will be called a term. In fact, a term is usually defined as a word or a multi-word expression that is used in specific domains with a specific meaning.

New relationships among concepts are continually created. In fact, Ausubel stated that the acquisition of new knowledge is dependent on what is already known [2]. To learn meaningfully, students must intentionally attempt to integrate new concepts with existing ones so that they can interact in the learner's knowledge structure. In this way, they get a more extensive network of knowledge and more retrieval paths. This theory has been successfully used in education during the last decades.

This inner flexibility of representations may cause that, from the source information, each individual acquires something that could be completely different. This has been called by some researchers as "a plague on attempts to educate and evaluate" [3]. In fact, some studies report that what learners actually learn and what they should understand as a consequence of instruction is often very different. Thus, instructors should be provided with tools that allow them to visualize the knowledge structure of their students. That is, instructors should be provided with the students' conceptual models, defining a conceptual model as a simplified representation of the concepts and relationships among them that a student keeps in his or her mind about an area of knowledge. In this way, teachers can identify where the main students' misconceptions are and, which concepts have already been assimilated. Each concept in the conceptual model must have associated an indication of the confidence in how well it is known by each student and the group of students.

The conceptual model can be graphically displayed as a concept map. In fact, the theory of Meaningful Learning is the fundamental pillar of concept maps that can be defined as powerful tools to visually represent the conceptual structure that someone has about an area of knowledge. Novak introduced them as a tool for students to freely organize their knowledge about a certain area [1].

This representation takes the form of a graph or a diagram that shows the concepts and the connections between them that students have. Three main basic elements can be identified in a concept map: the concepts that are represented in the graph as the nodes (in a concept map, each concept only appears once); the links that join two

This paper is based on "Automatic Generation of Students' Conceptual Models underpinned by Free-Text Adaptive Computer Assisted Assessment" by Diana Pérez-Marín, Enrique Alfonseca, Manuel Freire, Pilar Rodríguez, José-María Guirao and Antonio Moreno-Sandoval which appeared in the Proceedings of the 6th IEEE International Conference on Advanced Learning Technologies (ICALT), Kerkrade, The Netherlands, July 2006. © 2007 IEEE.

This work has been sponsored by the Spanish Ministry of Science and Technology, project number TIN2004-03140.

© 2007 ACADEMY PUBLISHER

related nodes (a node can be related to one or more nodes. Arrow symbols are used at the end of each link to describe the direction of each relationship); and, the propositions that are the basic units of meaning. They are created from the composition of the labels of the concepts and the label of their link that indicates the type of relationship between these nodes.

In this paper, an automatic procedure to automatically generate the conceptual model not only for one student but for the whole class from the free-text answers provided to an automatic and adaptive free-text Computer Assisted Assessment (CAA) system called Willow [4] is presented. It is based on the procedure described in [5], but extended with a better Term Identification technique based on Machine Learning [6] and, an enhanced visual representation of the conceptual model as a concept map using the COMOV system. Furthermore, a new metric to evaluate the confidence that a student knows each term in his or her conceptual model is proposed. We believe that in the automatic scoring of free-text answers, it is necessary metrics to evaluate how confidently can be stated that a student knows a certain concept.

The paper is organized as follows: Section II introduces the theoretical framework background and related work to the approach presented; Section III defines the conceptual model; Section IV describes the procedure of the generation of the conceptual model of a student and the whole class focusing on the proposed metric to evaluate students' knowledge of the concepts; Section V focuses on the visual representation of the conceptual model; and, finally, Section VI gathers the main conclusions drawn from this work and the most promising lines of future work.

II. RELATED WORK

Web-based distance education has rapidly become very popular. As well as in traditional teaching, it is crucial to determine for each course what to teach, how to teach and how to ensure learners' mastery of the material. Good human instructors can intuitively make these determinations, whereas computers must be programmed as Intelligent Tutoring Systems (ITSs) [7] or Adaptive Educational Hypermedia Systems (AEHs) [8].

Students may feel disoriented without the support of a tutor. The regular monitoring of the students' behavior can reduce these problems. Besides, educators demand to be able to have an overview of the performance of their students, monitor discussions, cluster learner groups based on certain patterns of behaviour/performance, identify tendencies in different groups and discover common misconceptions (not only for traditional students but also for distance ones). Hence, the instructors should be provided with tools to keep track of the students' models and so, to be able to prevent or overcome potential conflicts.

Concept maps are particularly useful for representing the networks of concepts in students' minds. They can be used to point out any conceptual misconceptions the person may have concerning the knowledge structure. This explicit evaluation of knowledge and subsequent recognition of misconceptions allows for finely targeted remediation. Furthermore, since concept maps are visual images they tend to be more easily remembered than text. More about concept maps and their underlying theory can be found in [9].

Hence, it could be concluded that concept maps should be the common knowledge representation tool today and they have extensively used in the classrooms. However, it is not the case. It could be due to the fact that they are time consuming to learn how to create them and difficult to manage in paper [10]. To solve this problem and thanks to the generalization of computers in education, many computer applications have been developed to support the creation and maintenance of concept maps. Automated tools can improve visual appearance and consistency. They also facilitate the display and revision of large and complex maps through functionalities such as zooming and automatic redraw. In particular, programs such as CMapTools [11] or CMTool [12] are helping in the introduction of concept maps in the computer to make easier their design and management. There are also many educational systems that are underpinned by the Meaningful learning theory of Ausubel and use concept maps as supporting tool. Some of them are: ALE, DynMap+, E-TESTER, LEO and STyLE-OLM.

ALE is an adaptive and adaptable learning environment that provides individualized education. Its main goal is to foster meaningful and multidisciplinary learning. It also allows discovery learning by providing concept-based navigation. Moreover, it keeps a model of the students with information about their learning style to adjust the navigation possibilities to them and supports coaching by relying on case-based reasoning. The student's model always reflects the current state of the student's progress to give the most suitable recommendations based on the student's learning style, preferences and knowledge stored in the model. The metric used to measure the confidence that a concept is known is to let the teacher indicate if a concept is already known or not [13].

DynMap+ is a graphical tool to display the student model as a concept map. Students introduce the concept map in the computer using an editor. DynMap+ can show models not only of individuals but also of groups. Both are overlay models that can be shown to students and instructors. The purpose of showing it to instructors is to provide them with a view of the knowledge and evolution of the students and, of showing the map to students is to foster reflective thinking about their own learning. No numerical metric is used to evaluate the confidence that a student knows a certain concept. It is rather displayed according to the size and the type of lines of the nodes of the concept map. In particular, from the comparison of the concept map created by the student and the domain model concept map, the size of the nodes (smaller when they are more unrelated) and the type of line (discontinuous when the content of this node is still incomplete) is determined. See Figure 1 (left) for a snapshot of the system [14].

© 2007 ACADEMY PUBLISHER

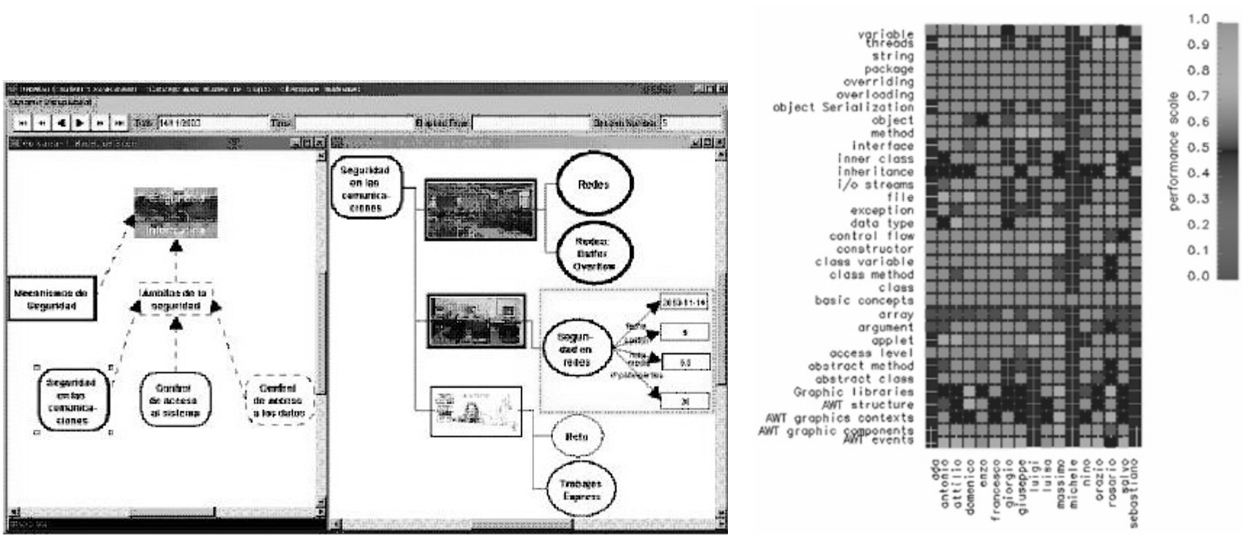

Figure 1. On the left, a snapshot of the Dynmap system and on the right, a snapshot of the CourseVis system.

E-TESTER is a computer-based system that identifies the main concepts in a text and generates questions from these concepts such as “What is *xxx*?” or “Explain *yyy*”. Next, it waits for the student answers in free-text to compare them with the e-learning content that the system has and treats as model answer. The comparison is based on the free-text scoring system Markit [15]. The difference is that in E-TESTER the process is simplified as it is only focused on counting the frequencies of the identified concepts in the student’s answer and in the model answers (i.e. the metric to measure the confidence that a student knows a certain concept is based on the frequency of use of this term by the student). In fact, its feedback is not a numerical score but a visual representation of the student strong and weak points [16].

LEO is a system based on the Ausubel’s Meaningful learning theory that provides students with a graphical schema of the course, links to instructional content and a visual representation of their progress. LEO is part of a the “CMapTools” software suite that allows experts to structure knowledge of a domain as a concept map. Like DynMap+, no numerical metric is given to measure the confidence value that a concept is known. Instead a color code is used. In particular, topic nodes have color codings at the left to indicate student progress through the course of instruction and the links between topic nodes convey prerequisite relationships among topics. Possible status for a topic node are: completed, ready, not ready and current. As the student works through the course, the model is being updated so that the next time that the student logs in the system s/he can see the updated map [17].

STyLE-OLM is a diagnostic tool integrated in the STyLE educational system that interactively builds the student model through a dialogue based on conceptual graphs between the student and the system. Its main goal is to engage students in reflective activities. The student model is an overlay of the domain model which incorporates the student’s beliefs and misconceptions formalized with Prolog rules. Beliefs can be correct (supported by the domain ontology), incomplete (facts from the domain ontology that the student does not believe and can be elicited by the system or stated by the student with an “I do not know answer” to a system’s question) or erroneous (not supported by the domain ontology), and are open for inspection and discussion. The student model can be visually depicted with **CourseVis** as can be seen in Figure 1 (right). It shows a cognitive matrix in which the students are mapped onto the x-axis and the concepts of the course are mapped onto the y-axis. The performance values (i.e. the numerical metric used to measure the confidence that a student knows a concept) are mapped onto the color of the square corresponding to a student and a concept. This matrix is shown to instructors so that they can detect problematic topics, struggle students by comparing columns and row or analyze the performance of a particular student on a topic [18].

These systems are only a sample. There are many other educational systems that are underpinned by some kind of conceptual model. However, none of them attempts to fully generate the conceptual model from the students’ free-text answers. Regarding free-text answers scorers is still regarded by many as the Holy Grail of Computer Assisted Assessment (CAA). However, the advances in Natural Language Processing (NLP), Machine Learning and Neural Network techniques, the lack of time to give students appropriate feedback (despite the general assumption of its importance) and the conviction that Multiple Choice Questions (MCQs) cannot be the only assessment method are favoring a change in this situation. In fact, there are currently more than fifteen systems that tackle this problem with the core idea of comparing the student answer with one or more reference texts and the more similar they are, the higher the score the answer achieves.

Some of these free-text CAA systems are: the **Project**

© 2007 ACADEMY PUBLISHER

Essay Grader (PEG) [19] that is the oldest free-text CAA system. It was originally based on analysis of syntactic features such as number of words, length of the words, etc. that served to analyze more abstract information such as the use of vocabulary, language proficiency, etc.; the **The Intelligent Essay Assessor (IEA)** [20] that focuses on the content and it is based on Latent Semantic Analysis (LSA); **E-rater** [21] that measures the organization, the sentence structure and the content of the answer by using a hybrid approach that combines NLP with statistical techniques; **Automark** [22] that is based on Information Extraction techniques that perform an intelligent search of free text responses according to predefined computerized mark scheme answers; and, **MRW** [23] that provides a new approach to the field based on transforming the texts to score in semantic networks and next, by comparing the semantic networks instead of the plain text.

Up to our knowledge, none of these systems takes currently into account a student's model to adapt the assessment. Thus, free-text Adaptive Computer Assisted Assessment (ACAA) is the natural evolution of free-text CAA systems in which not only NLP techniques but also Adaptive Hypermedia techniques are applied to achieve a new way of formative assessment that fits better the needs of the students as the content, order and references of the questions are chosen according to the information known by the system about the student [4].

III. Definition of the conceptual model

A conceptual model can be defined as a simplified representation of the concepts and relationships among them that each student keeps in his or her mind about an area of knowledge in a certain instant. Conceptual models have been extensively used for many different applications such as summative and formative assessment, knowledge elicitation and organization, etc. In this work, two types of conceptual models are addressed: those representing the knowledge of each student, and those representing the averaged knowledge for the whole class.

A. Types of concepts

A hierarchical structure of knowledge is considered, according to which, not all concepts in the model have the same relevancy. In fact, the higher they are in the hierarchy, the more important they are. Three different types of concepts have been distinguished:

- **Basic-concepts (BCs)**: Specific terms relevant for one or more topics. They are in the the lowest level in the hierarchy, as they refer to individual instances (i.e. what Novak considers concepts). For example, *blanket*, *semaphore* or *process*. They are automatically extracted from the free-text students' answers as explained in Section IV.
- **Topic-concepts (TCs)**: Main issues inside an area of knowledge. They are an intermediate level in the hierarchy as they group several BCs and belong to a certain area-of-knowledge concept. For example, *concurrency* is a TC that comprises BCs such as *semaphore* and *process* but not *blanket*. It belongs to the area-of-knowledge concept *operating system*. TCs are extracted from the names of the lessons of the agenda of the course as provided by the instructors.
- **Area-of-knowledge-concepts (ACs)**: Main domains of knowledge that contains all the rest of the concepts. That is, they are the higher level concepts as they refer to groups of several TCs. For instance, *operating system* is an AC that comprises topic-concepts such as *concurrency* or *scheduling*. For each conceptual model, only one AC is allowed and it corresponds to the name of the course to model as given by the instructors.

Each concept, irrespectively of its type, has a confidence-value (CV) that reflects how well it is understood at any time. It is always between 0 and 1. A lower value means that the student does not know the concept as s/he does not use it, while a higher value means that the student confidently uses that concept. This CV is automatically updated as the student keeps answering questions according to the set of metrics proposed in this work and described in Section IV. The CV of a TC is calculated as the mean value of the CVs of the BCs that groups and the CV of an AC is calculated from the CVs of its TCs. Thus, just by looking if the AC has a high confidence value, it can be seen how well the whole course has been understood.

B. Types of relationships between concepts

Regarding the relationships between these concepts, three types of links have been distinguished according to the type of concepts that they relate:

- **Type 1, between ACs and TCs**: A topic-concept belongs to one area-of-knowledge-concept. For example, the TC *concurrency* belongs to the AC *operating system*. Type 1 links are extracted from the organization of the course provided by the instructors (i.e. which lessons corresponds to each course). A TC can only belong to one AC.
- **Type 2, between TC and BC**: A basic-concept belongs at least to one topic-concept. It can also belong to several topic-concepts. For example, the BC *semaphore* belongs to the TC *concurrency* while the BC *process* belongs to the TCs *concurrency* and *scheduling*. These relationships are important because they give us information about how the basic-concepts are grouped in topic-concepts. Moreover, for each BC that belongs to different TCs, the student's ability to deal with the BC in the different contexts provided by the TCs. TCs are not linked among them, as the relationships between the topics are already captured by the type 3 links. Type 2 links are extracted from the database of the course, in particular, from the relationships between the topics and the BCs that belong to each topic.

© 2007 ACADEMY PUBLISHER

- **Type 3, between two BCs**: A basic-concept can be related to one or more basic-concepts. For example, the BC *process* is related to the BCs *program* and *thread* as "a process is an instance of a program" and "a process can contain one or more threads". These links are very important as they reflect how BCs are related in the student's cognitive structure as are extracted from the students' answers.

Type 1 and 2 links are equal to all students as they are extracted from the structure of the course that is common to all of them, whereas type 3 links are specific to each student as they are extracted from their answers. It is also important to mention that each link has associated one or more linking words that join the concepts in the extremes of the link and form propositions.

The linking words for type 1 and type 2 links have been fixed as "talks about" (from the higher concept in the hierarchy to a lower concept) or "'belongs to" (from the lower concept in the hierarchy to a lower concept). These linking words have been chosen as they serve to structure the knowledge and thus, capture the essence of these type of links. For example, *operating system* "talks about" *concurrency* or the other way around, *concurrency* "belongs to" *operating system*. In the case of type 3 links, the linking words are directly extracted from the text in which the link has been found. For example, for the BCs *program* and *software*, the proposition *program* "is a kind of" *software* can be found. It does not only indicate that the student knows the BCs *software* and *program*, but also that s/he has meaningfully learnt the BC *program* as s/he knows how to link it it with *software*.

IV. Generation of the Conceptual Model

The procedure consists of three main steps that are to: find the concepts, estimate the confidence-value that each concept found is known according to the metrics proposed and, finally find the relationships between the concepts. These steps are described below.

A. Finding the concepts

In order to build the conceptual model, a necessary first step is to extract all basic concepts as the labels of the concepts (i.e. terms) used by the instructors in the references. Besides, the AC is defined as the name of the area-of-knowledge and the TCs as the names of the topics or lessons of the area-of-knowledge.

A term is usually defined as a word or a multi-word expression that is used in specific domains with a specific meaning. In particular, it is proposed to consider the term as a single word (unigram), a sequence of two words (bigram) or three words (trigram). Thus, each n-gram found in the text, with n varying from 1 to 3, can be classified as either being a term (class 1) or not (class -1). The focus is on nominal terms (i.e. nouns or multi-word noun phrases). Thus, from now on "term" will be used to refer to "nominal term".

Term Extraction is an important problem in the NLP area. It has received the attention of many researchers that have proposed several solutions to tackle it. Most of them rely on the analysis of large collection of domain-specific texts and compare them to general-purpose text, in order to find domain-specific regularities that indicate that a particular word or multi-word expression is a relevant term in that domain.

In education, the traditional approach has been to ask human experts to choose these terms [24]. However, there are several critics to this approach, as leaving the decision to humans makes it subjective and two humans tend not to agree completely. However, up to our knowledge, none of the existing term extraction techniques have been applied to automatically identify the main terms of an area-of-knowledge to serve as basis of the generation of a student's conceptual model. It would be interesting not only to make the procedure more objective but to free the instructors of the additional task of having to identify these terms by themselves.

Therefore, we propose a new module able to use term extraction techniques to automatically identify the terms (main concepts) of a particular knowledge field as they appear in a set of references [6]. Our approach is based on the technique described by [25]. That is, to use the C4.5 algorithm to learn a decision tree. Due to its statistical nature, this algorithm has the advantage of being equally applicable to different languages such as Spanish and English.

The decision tree has been trained with a set of references (i.e. the domain-specific corpus), whereas the generic corpus can be as general as the British National Corpus or to treat about the same area that the domain-specific corpus but in a more general way. Besides, it is advisable for the learning phase to choose the samples to produce a balanced distribution of classes (50% terms and 50% non-terms). Moreover, the features considered as attributes should be at least: the relative frequency of appearance of the term in a corpus of students answers with respect to its frequency in the generic corpus and the sequence of part-of-speech tags of the words (e.g. noun, verb, adjective, etc.). The reason of choosing these features was that they are related to the nature of which a term is:

- The relative frequency is important because terms tend to be specific to a certain knowledge field. Thus, words with a relative frequency (frequency in the specific corpus / frequency in the generic corpus) lower than one should be discarded as they are too common.
- The part-of-speech (POS) is relevant because it is only considered noun or a simple multi-word noun phrase as a possible term. In most of the cases, the syntactic structure of noun phrases is not as complex as that of a clause or a sentence, so it can be characterized by using regular expressions on the POS tags. For instance, the sequence of tags "determiner+noun+adjective" covers noun phrases such as "the operating system". Thus, if a word is a finite-tense verb it will probably not be part of a

© 2007 ACADEMY PUBLISHER

nominal term.

- Moreover, it has been observed that usually the considered terms can be represented by the following regular expression: NC* NP* ADJ* PREP* NC2* NP2* (zero or more common names, proper names, adjectives, prepositions, more common names and more proper names). Thus, each n-gram extracted from the corpus are matched to the previous regular expression, giving to each of the POS tags a weight equal to the number of words belonging to that class. Later, the weights are normalized so that they all add up one.

It is also important to highlight that the result given by the algorithm can be modified by the instructor. S/he is allowed to remove or add terms.

B. Estimating the CV of each student's concepts for each student's conceptual model

As it has been seen in Section II, several approaches have been tried to estimate the Confidence-Value (CV) of the concepts used by the students. These approaches can be as simple as just to find out if the student has accessed the learning object to mark it as seen or not seen; use some information extracted from the student's performance such as the frequency of use of the term that labels the concept in his or her answer [16] or how s/he has solved some quizzes or exercises [26], [27]; or, as complex as to the calculation of sophisticated bayesian belief values [28]. However, there is currently a lack of a standard metric for estimating the CV of a BC that can be used irrespectively of the ad-hoc implementation of a particular system. Therefore, a new metric to estimate the CV of a concept taking advantage of the existing approaches is proposed in this work.

The metric to estimate the CV of each student's concept has to include information about the student's performance (e.g. the score achieved in an answer or the use of terms) and information about the reference knowledge such as the use of terms in the correct answers provided by the instructors or the e-learning reference content [16]. It is because, it is always necessary some kind of reference knowledge to allow the automatic assessment of free-text answers by comparing the answer provided to this reference knowledge. In particular, it is assumed that the higher the score of this comparison as they are more similar (i.e. students should use the terms in the answers as instructors do in the correct answers), the better the knowledge the student has about the topic under assessment. Moreover, given the hierarchical structure of knowledge proposed, it can be said that the higher the score of this comparison, the higher the confidence that the student knows the terms (labels of the concepts) used in the answer as they belong to the topics of the area-of-knowledge under assessment. To sum up, the metric has to: include information about the reference knowledge, include information about the student's performance when answering questions and, take into account that the student's answer should be similar to the reference answer if it is correct.

To give a special relevancy of each of these requisites, it has been agreed to discern these two metrics: $ScoreConfidence$ (SC) and $RateConfidence$ (RC) to be used together (as shown in Equation 10) in order to estimate the CV that an individual i knows a certain concept c labeled by a term t taking as reference knowledge, the answers provided by a set of questions Q_i in the language L.

In particular, SC is more focused on the first and second requisites as it includes the score that the free-text ACAA system gives to the answer, so that the higher the score, the higher the CV of c labeled by t as the student is correctly using t. As will be seen in Equation 8, it is the mean of the weighted scores for the set of questions whose references contain t (i.e. it fulfils the second requisite). Besides, as the weight assigned to each score is calculated as the mean between the frequency of t in the references of the question and all the references, it also fulfils the first requisite.

RC is more focused on the third requisite. That is, it is more related to the comparison of the frequency of t in the answer provided by i and the frequency of t in the correct answers taken as references. In fact, as can be seen in Equation 9, it is calculated as the mean of the ratio between the frequency of t in the answers provided by i and the references of all the questions in the area to assess.

Once the metrics have been justified and an intuitive definition has been provided for each of them, it is convenient to formalize these ideas according the following mathematical notation in order to provide a more precise definition. Please, note that when lowercase is used, a particular element is referred while uppercase refers to groups of elements. Let the variables be:

- W, that is the set of words of a language defined as:

$$W = \{w | w \ is \ a \ word \ of \ L\} \tag{1}$$

- P, that is the set of phrases defined as:

$$P = \{x | x \in \bigcup_{n=1}^{\infty} W^n \wedge x \ is \ a \ noun \ phrase\} \tag{2}$$

- S, that is the set of possible sentences defined as:

$$S = \{x | x \in \bigcup_{n=1}^{\infty} W^n \wedge x \ is \ a \ sentence\} \tag{3}$$

- Q_A, that is the set of questions to assess an area-of-knowledge A.
- I, that is the set of individuals whose knowledge in A is evaluated using Q_A.
- Q_i, that is the set of questions asked to i.
- T_A, that is the set of terms extracted from A.

Let the functions be:

- $score$, that assigns the numerical mark given by the free-text ACAA system to the answer provided by I to the question q $\in Q_A$:

$$score : I \times Q_A \longrightarrow \Re \tag{4}$$

© 2007 ACADEMY PUBLISHER

- $frequency$, of $t \in T_A$ for a set of sentences $s_i \in S$:

$$f : T_A \times \bigcup_{n=1}^{\infty} S^n \longrightarrow N \tag{5}$$

- $references$, for Q_A:

$$refs : \wp(Q_A) \backslash \emptyset \longrightarrow S \tag{6}$$

- $answer$, provided by $i \in I$ to the question $q \in Q_A$:

$$answer : Q_A \longrightarrow \bigcup_{n=1}^{\infty} S^n \tag{7}$$

From these definitions, Equation 8 gives the mathematical formulation of the first metric:

$$SC(t,i,Q_i) = \frac{\sum_{q \in Q_i} score(i,q) \times f(t, refs(\{q\}))}{f(t, refs(Q_i))} \tag{8}$$

The possible values of SC are the same that the $score$ function as it is its weighted mean. Equation 9 gives the mathematical formulation of the second metric:

$$RC(t,i,Q_i) = \frac{\sum_{q \in Q_i \wedge f(t,refs(q)) \neq 0} \frac{f(t,answer(i,q))}{f(t,refs(q))}}{||Q_i||} \tag{9}$$

The possible values of RC range from 0 up to a positive K that depends on the language that weights the maximum number of words that could contain its longest sentence. As can be seen, the range of possible values of SC and RC are not compatible, and thus, in order to make possible their combination, they have to be normalized. It is done by dividing by the maximum value of the range so that both metrics are scaled to the range 0 (minimum confidence in that the concept is known) up to 1 (maximum confidence in that the concept is known).

Therefore, the metric for CV can be defined as the function that assigns a number from 0 up to 1 that indicates the confidence that the free-text ACAA system has that a certain $i \in I$ knows $t \in T_A$ according to the answers provided and the references of a set of questions $Q_i \in Q_A$. The weight given to SC and RC depends on which requisite of the metric is considered most relevant. The weight can be initially fixed (by default) to 50% so that both metrics have the same relevancy. Equation 10 is its mathematical formulation:

$$CVScore_{\beta}(t,i,Q_i) = SC_n(t,i,Q_i)\dot{\beta} + RC_n(t,i,Q_i)\dot{(}1-\beta) \tag{10}$$

By using this formula, each term is assigned a CV. Moreover, the BC labeled by this term is assigned a CV. In the case of TCs, the underlying idea is that since BCs belong to one or more TCs, a TC has been understood if all its BCs have been understood. Hence, the CV of each TC is calculated as the mean value of the CVs of all the BCs related to it. Similarly, once each TC has been assigned a CV, the AC confidence-value is calculated as the mean value of the CVs of the TCs related to it.

C. Estimating the CV of the student's concepts for the whole class conceptual model

The same ideas are used for the whole class conceptual model, with the difference that the values are averaged for all the answers given by all the students of the class. In this way, what is modeled is not the particular use of a BC, TC or AC by a student but by the whole class. Equation 11 shows the mathematical formulation of the first metric for a set I of students:

$$SCGroup(t,I,Q_I) = \sum_{i \in I} \frac{SC(t,i,Q_i)}{||I||} \tag{11}$$

Similarly, Equation 12 is the mathematical formulation of the second metric, and Equation 13 is the mathematical formulation of the CV to be assigned to each BC of the whole class conceptual model:

$$RCGroup(t,I,Q_I) = \sum_{i \in I} \frac{RC(t,i,Q_i)}{||I||} \tag{12}$$

$$CVScoreGroup(t,I,Q_I) = \sum_{i \in I} \frac{CVScore(t,i,Q_i)}{||I||} \tag{13}$$

TCs and ACs confidence values for the group conceptual model are calculated as in the particular student conceptual model but from the values achieve with Equation 13 so that the results are common to all the students.

D. Finding the relationships between the concepts

The content of the conceptual model is not only the concepts and their CVs but also, and very importantly, the relationships between these concepts. Type 1 links (AC-TC) are fixed according to the information provided by the instructor. In fact, they are created by connecting the AC with each TC. Thus, they are the same for all the students (although it is important to note that the CVs of the concepts that join are different).

Type 2 links (TC-BC) are created once the BCs are extracted using the automatic term identification module and transformed to their canonical form (plural to singular, feminine to masculine, etc.). The idea is that in the term that labels a BC has been found in a reference of a question q belonging to a certain topic, then this BC has to be linked to the name of the topic, that is, its TC. Hence, as above, they are the same for all the students (although the CVs of the concepts to be related are different). It is also important to mention that a BC can belong to different TCs and thus, several type 2 links can be created connecting the BC to each TC.

Finally, type 3 links (BC-BC) are automatically extracted from the answers provided by the students. The procedure is as follows: find one BC and mark it as the first BC of the relationship; find another BC in the same sentence and mark it as the second BC of the relationship; and, finally, extract the words between the first and the second BC and mark them as the linking words of the relationship.

© 2007 ACADEMY PUBLISHER

V. VISUALIZATION OF THE CONCEPTUAL MODEL

As it has previously stated, concept maps have been widely used for all aspects of education and they are considered as one of the best representations to intuitively show how concepts are interrelated in people's minds and where the misconceptions and lack of previous concepts are. This has been the motivation to represent the generated conceptual model as a concept map composed by nodes (each node represents a concept) and links between them. A spider-like organization of the map has been chosen, as it is one of the most suitable formats for the hierarchy of concepts proposed with the AC in the center of the map, the TCs linked to it in their surroundings and the BCs adjacently linked to their related TCs and BCs.

Originally, the CLOVER [29] tool was used to visualize the conceptual model as a concept map. Currently, a new viewer called IOV (Integrated On-line Viewer) has been developed. It shows a simplified representation of the conceptual model as a concept map with the aim of making it more clear to the users when the number of concepts of the course is below a certain threshold. Three main modifications have been implemented:

- **The clustering of concepts has been deactivated**: It is because users sometimes complaint that the CLOVER representation was too complex to understand. Hence, as the hierarchy of the knowledge is already implicitly represented by the use of ACs, TCs and BCs, no additional grouping is seen as necessary and, to make this representation more similar to the original concept map proposed by Novak, in the new viewer a node always represents a concept and not a group of concepts.
- **The color schema has been changed**: To use two different color codes for the background and foreground colors was confusing to some users. Thus, the color schema is modified so that only the background color remains and it serves to represent the CV. Besides, it is not so strict for the ranges exposed (i.e. only red, yellow or green), but the whole degradation from utter red (CV=0) up to utter green (CV=1) passing from lighter red, orange, lighter orange, yellow, stronger yellow, light green and stronger green. The foreground color is no longer necessary, as the the type of node is indicated by the size and place in the concept map. The AC is bigger and it is always at the center, the TCs are medium-size and are placed in the second radial line while the BCs are smaller and are placed in the next radials lines (red BCs nearest to the TCs and greener BCs furthest to give a general impression of degradation from the borders of the concept maps).
- **The links have been reorganized**: An effort have been done to avoid crossings among links. It has been achieved by locating each TC as a different branch parting from the AC and calculating the space for each node according to the nodes that are related to it.

See Figure 2 for an example of a student's conceptual model represented as concept map with IOV. It can be seen that it is easy to discern if the student has successfully assimilated the concepts exposed in the lesson just by looking at the higher concept of the hierarchy (the AC). If it has a green foreground color, it means that the student is ready to continue learning another AC. Otherwise, some problems have appeared and they can be identified by looking at lower concepts in the hierarchy, initially TCs to see which ones are lacking and next, the BCs related to the non green TCs.

Additionally to the information provided by the concepts and its hierarchy, links are very useful to detect misconceptions and lack of relationships. The misconceptions are detected whenever there is a type 3 link between two BCs that should not be related and thus, teachers should evidence the erroneous of this link. The lack of type 3 links denotes that students may understand each isolated concept but they have not recognized that they are related and thus, teachers need to reinforce the link between them.

This representation in form of concept map is particularly interesting whenever a global view of a particular student or the whole class is pursued. Moreover, when the evolution through the course wants to be reviewed

VI. CONCLUSIONS AND FUTURE WORK

Naturally, teachers aim that students acquire certain meanings that are accepted in the context of the course and shared by a users' community. However, it has been reported in the literature how what the teachers try to transmit to the students and what students are actually able to understand is quite different. In order to bridge this gap and improve the quality of teaching, we have presented in this paper an extended version of the procedure presented in [5] to automatically generate a student's and a group of students' conceptual model from their answers to an automatic free-text scoring system and to give it as feedback to teachers and students represented as a concept map.

This approach unlike the ones reviewed in the related literature, do not ask students to draw the concept maps or to negotiate it, but automatically generate them, making the whole process transparent both to teachers and students. Furthermore, it is different from the one proposed in E-TESTER as it is not only focused on comparison of frequencies in the student's answer and the references, but also take into account additional information such as the automatic score provided by the free-text ACAA system and the links among the concepts.

A hierarchical structure of knowledge is proposed in which three type of concepts are distinguished: BCs, TCs and AC. Besides, there can be identified three types of links: BC-BC, BC-TC and TC-AC. A new metric has been proposed to evaluate the confidence that a student knows a certain concept taking into account not only the frequency of use of the terms in the students' answers and the references, but also the automatic score and several

© 2007 ACADEMY PUBLISHER

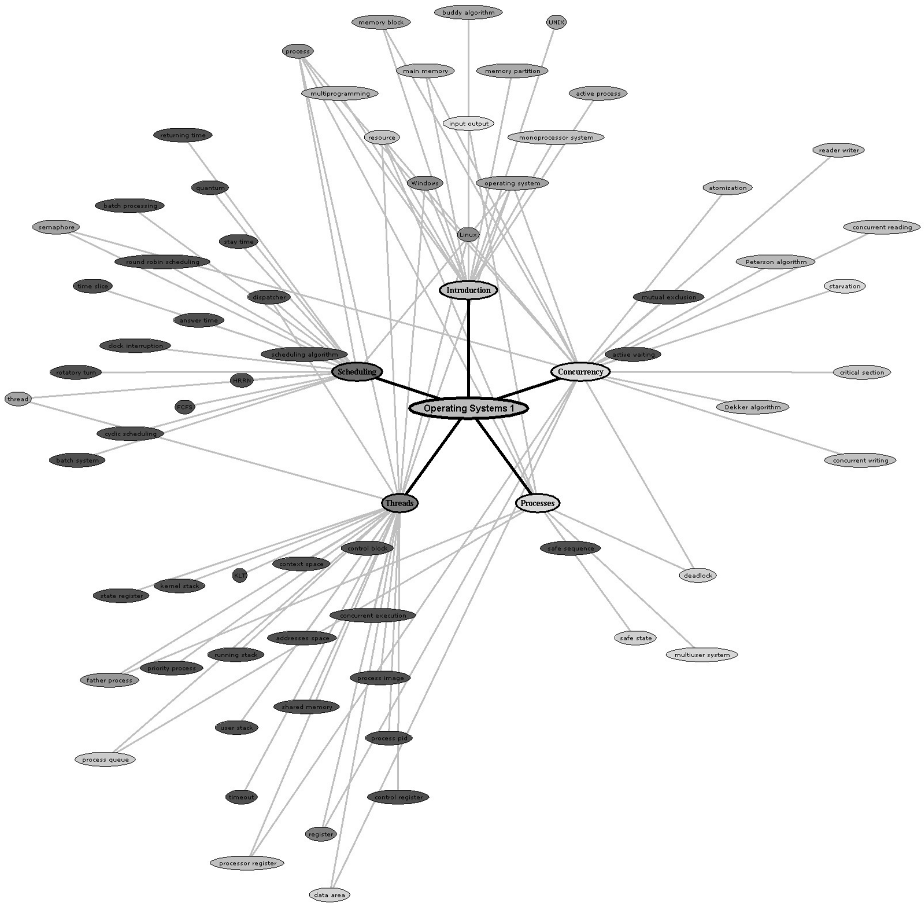

Figure 2. An example of concept map of a student that has studied a month of a course.

other ratios of comparison between the student's answer and the references.

No qualitative score has been assigned to the conceptual model as it is not intended its use as a new summative assessment tool. Our goal is not to keep the student model so that Willow can choose better the next question, to ask the students to draw concept maps to improve their reasoning process or to modify the score depending on the use of concepts. Our goal is to help teachers, to reveal them which concepts should be reviewed as the students are not understanding them (they failed to use them or use them incorrectly) and which ones have already been assimilated. Moreover, the conceptual model is not only particular to one student but it can be referred to the whole class. It cannot only be used in traditional courses but also and with a special interest in e-learning courses in which teachers are unable to directly talk to the students and the labor of finding their conflictive concepts is even more difficult.

Some lines of future work are:

- To give the possibility of generating the student or class conceptual model directly from students' answers in plain text without the necessity of asking the students of using Willow.
- To use the generated conceptual model to decide which should be the next question to ask the students according to the lack of knowledge or misconceptions detected in the previous answers provided to Willow.
- To represent the concept map in other formats such as diagram, bar chart, table, etc.
- To analyze the possibility of assigning a quantitative score to each concept map and, calculate the correlation between the scores achieved by the students in a real exam and the values associated to the generated concept maps.

ACKNOWLEDGMENT

We thank the teachers and students that collaborated with us in the experiments carried out.

REFERENCES

[1] J. Novak and D. Gowin, *Learning How to Learn*. Cambridge, U.K.: Cambridge University Press, 1984.
[2] D. Ausubel, J. Novak, and H. Hanesian, *Educational Psychology: a cognitive view, 2nd. ed.* New York: Holt, Reinhart and Winston, 1978.
[3] I. Sigel, Ed., *Development of mental representations: Theories and Applications*. New Jersey, U.S.A.: Lawrence Erlbaum Associates, 1999.
[4] D. Pérez-Marín, E. Alfonseca, and P. Rodríguez, "On the dynamic adaptation of computer assisted assessment of free-text answers," *In proceedings of the Adaptive Hypermedia Conference, LNCS 4018, Springer-Verlag,* 2006.

© 2007 ACADEMY PUBLISHER

[5] D. Pérez-Marín, E. Alfonseca, M. Freire, P. Rodríguez, J. Guirao, and A. Moreno-Sandoval, “Automatic generation of students’ conceptual models underpinned by free-text adaptive computer assisted assessment,” in *Proceedings of the IEEE International Conference on Advanced Learning Techniques (ICALT)*, 2006.

[6] D. Pérez-Marín, I. Pascual-Nieto, E. Alfonseca, and P. Rodríguez, “Automatic identification of terms for the generation of students concept maps,” in *In proceedings of the International Conference on Multimedia and Information Technologies for the Education (MICTE)*, 2006.

[7] V. Shute and L. A. Torreano, “Formative evaluation of an automated knowledge elicitation and organization tool,” *Authoring Tools for Advanced Technology Learning Environments: Toward Cost-Effective Adaptive, Interactive, and Intelligent Educational Software*, 2002.

[8] P. Brusilovsky, “Adaptive educational hypermedia: From generation to generation,” in *Proceedings of the 4th Hellenic Conference on Information and Communication Technologies in Education*, 2004.

[9] J. Novak and A. J. Canas, “The theory underlying concept maps and how to construct them,” Florida Institute for Human and Machine Cognition, Tech. Rep., 2006.

[10] L. Hsu, R. Edd, S. Hsieh, and R. Msn, “Concept maps as an assessment tool in a nursing course,” *Journal of Professional Nursing*, vol. 21, no. 3, pp. 141–149, 2005.

[11] A. Caas, D. Leake, and D. Wilson, “Managing, mapping and manipulating conceptual knowledge,” in *AAAI Workshop Technical Report WS-99-10: Exploring the Synergies of Knowledge Management and Case-Based Reasoning*. California, U.S.A: AAAI Press, 1999.

[12] F. Rocha and E. Favero, “Cmtool: A supporting tool for conceptual map analysis,” in *Proceedings of World Congress on Engineering and Technology Education*, Santos, Brazil, 2004.

[13] M. Kravcik and M. Specht, “Flexible navigation support in the winds learning environment for architecture and design,” *Proc. of the AH 2004 Conference*, 2004.

[14] U. Rueda, M. Larrañaga, A. Arruarte, and J. Elorriaga, “Modelado de grupos en actividades de aprendizaje basado en mapas conceptuales,” *Revista Iberoamericana de Inteligencia Artificial*, vol. 8, no. 24, pp. 131–140, 2004.

[15] R. Williams and H. Dreher, “Automatically grading essays with markit,” in *Proceedings of Informing Science 2004 Conference*, Rockhampton, Queensland, Australia, 2004.

[16] C. Guetl, H. Dreher, and R. Williams, “E-tester: A computer-based tool for auto-generated question and answer assessment,” *E-Learn*, 2005.

[17] J. W. Coffey, “Leo: A concept map based course visualization tool for instructors and students,” *KNOWLEDGE AND INFORMATION VISUALIZATION: SEARCHING FOR SYNERGIES*, vol. 3426, pp. 285–301, 2005.

[18] V. Dimitrova, “Style-olm: Interactive open learner modelling,” *International Journal of Artificial Intelligence in Education*, vol. 13, no. 1, pp. 35–78, 2003.

[19] E. Page, “The imminence of grading essays by computer,” *Phi Delta Kappan*, vol. 47, no. 1, pp. 238–243, 1966.

[20] P. Foltz, D. Laham, and T. Landauer, “The intelligent essay assessor: Applications to educational technology,” *Interactive Multimedia Electronic Journal of Computer-Enhanced Learning*, vol. 1, no. 2, 1999.

[21] J. Burstein, K. Kukich, S. Wolff, C. Lu, M. Chodorow, L. Bradenharder, and M. D. Harris, “Automated scoring using a hybrid feature identification technique,” in *Proceedings of the Annual Meeting of the Association of Computational Linguistics*, 1998.

[22] T. Mitchell, T. Russell, P. Broomhead, and N. Aldridge, “Towards robust computerised marking of free-text responses,” in *Proceedings of the 6th Computer Assisted Assessment Conference*, 2002.

[23] R. Lutticke, “Graphic and nlp based assessment of knowledge about semantic networks,” in *Proceedings of the Artifical Intelligence in Education (AIED) conference*, 2005.

[24] M. Ruiz-Primo, “Examining concept maps as an assessment tool,” in *Concept Maps: Theory, Methodology, Technology. Proceedings of the First International Conference on Concept Mapping*, Pamplona, Spain, June 2004.

[25] A. Ballester, A. Martín-Municio, F. Pardos, J. Porta-Zamorano, R. Ruiz-Urena, and F. Sánchez-León, “Combining statistics on n-grams for automatic term recognition,” in *In Proceedings of the Language Resources and Evaluation Conference (LREC)*, 2002.

[26] R. Mazza and V. Dimitrova, “Coursevis: Externalising student information to facilitate instructors in distance learning,” in *Proceedings of 11th International Conference on Artificial Intelligence in Education (AIED03)*, 2003.

[27] M. Muehlenbrock, S. Winterstein, E. Andres, and A. Meier, “Learner modeling in iclass,” in *Proceedings of the World Conference on Educational Multimedia, Hypermedia, and Telecommunications EdMedia*, Montreal, Canada, 2005.

[28] J. Zapata-Rivera and J. Greer, “Externalising learner modelling representations,” *Proceedings of Workshop on External Representations of AIED: Multiple Forms and Multiple Roles*, pp. 71–76, 2001.

[29] M. Freire and P. Rodríguez, “A graph-based interface to complex hypermedia structure visualization,” in *Proceedings of the Working Conference on Advanced Visual Interfaces (AVI), ACM Press*, 2004, pp. 163–166.

Diana Pérez-Marín is currently a lecturer and a Ph.D. candidate at Universidad Autónoma de Madrid (UAM), Spain. She received her MS and BS degrees in Computer Science from the UAM in 2004 and 2001, respectively. Her research interests are focused on the integration of Adaptive Hypermedia (AH) and Natural Language Processing (NLP) techniques. In particular, in the possibility of automatically generating inspectable open students’ conceptual models from the students’ answers to adaptive free-text Computer Assisted Assessment systems.

Dr. Enrique Alfonseca is a lecturer at the Computer Science and Engineering Department at the Universidad Autnoma de Madrid (UAM). He received his Ph.D. in Computer Science from UAM in 2003. His research verses on the application of NLP techniques for information extraction, ontology learning and hypermedia generation.

Dr. Pilar Rodríguez is a lecturer at the Computer Science and Engineering Department at the UAM. She received her Ph.D. from the Universidad Complutense de Madrid in 1990. Her Ph.D. thesis dealt with computational linguistics (Machine Translation). She has led and participated in numerous research projects (national and EU-funded) on web-based education.

Ismael Pascual-Nieto is a grant-holder and Ph.D. student at the UAM. He got his BS in Computer Science in the UAM in 2005. His research interests are: Natural Language Processing (NLP) techniques such as Information Extraction and Machine Translation, Knowledge Representation and the integration of AH and NLP techniques. Currently he is working in his MS which is focused on Translation Memories and is expected to be presented in 2007.

© 2007 ACADEMY PUBLISHER