Development of a Web System utilizing Bayesian networks for teaching English in the "Zoila Alvarado de Jaramillo" primary school

Katherine Ramón Campoverde[a], Luis Jácome Galarza[b], Iván Siguencia Larreátegui[c]

[a] System Engineering, Universidad Nacional de Loja (UNL), Loja, Ecuador
kiramonc@unl.edu.ec

[b] Instructor of Systems Engineering , Department of Energy, Industry, and the Non-renewable Natural Resources, of the UNL. Informational Systems and Computation Engineer. Master in Telematics. Master student in Computational Engineering and Intelligent Systems.
lrjacome@unl.edu.ec

[c] Instructor at the UNL. Obtained his BS and MS at the New Jersey Institute of Technology (NJIT).
Ivansl1971@gmail.com

Abstract. This paper proposes a method of English teaching through formative assessment. The proposal is based on a web application aimed at children from 5-6 years to teach words with computerized adaptive test using Bayesian networks as a method of adaptation, for which the library is used OpenMarkov. It is evaluated: the child's ability to hear and pronounce words. For pronunciation, speech recognition technique has been implemented with the help of the library is used Annyang. The validation tests of web application were performed in "Zoila Alvarado de Jaramillo" at Loja city.

Key Words: Bayesian network, student modelling, intelligent tutoring system, computerized adaptive test, OpenMarkov, speech recognition, Annyang.

1 Introduction

The teaching-learning process is complex, reason why there has been wide development of platforms for this purpose. Many educational systems have also been developed [1]. Information technology has demonstrated that it's possible to provide the student individualized instruction including privacy. In the context, Intelligent Tutoring Systems have been developed in connection with Bayesian networks. Bayesian networks are statistical tools developed in the context of artificial intelligence. These tools are focused to probabilistic inference and in the education realm are used to model the uncertainty related to the student and his level of knowledge [2]. This article's structure follow the following format:

In section 2 we introduce the adaptive informative tests, this section was divided in two parts, in section 2.1 the primary elements are indicated necessary for the construction of an adaptive informative test and in section 2.2 we show the general algorithm for the functioning of an adaptive informative test. Section 3 shows how the adaptive test has been defined for this Project. This section was further divided in two parts. In section 3.1 we show the structure utilized in the Bayesian network. Section 3.2 explains how the elements are defined for the adaptive informative test. Section 4 contains information on the implementation of this Project. Section 5 contains the results obtained after test were done with a group of students of the primary school Zoila Alvarado de Jaramillo located in the city of Loja. Finally, in section we discuss the conclusions obtained in this study.

2 Computer adaptive testing.

The use of tests for evaluation is a widely used technique in education as shown in [3]. The traditional methods of design and administration of tests oriented to groups have the advantage to be less costly in time and resources compared to individual tests. Additionally, all the tests are in equal conditions. This fact can also bring undesirable consequences like boredom of high achieving students or frustration of low advantaged students. In the 80's there was a surge of the so-called adaptive informative tests (CAT) that basically are tests administered by computers where the presentation of each question and the decision of finalizing the test are taken in a dynamic style base in students response and the estimation of their level of knowledge [3].

2.1 Elements of computer adaptive testing

The basic elements of a CAT are:

Model item response. This model describes how the subject responds to each item according to his level of knowledge. The measure should be invariant with respect to the subject which the test is applied to.

Theory response for each item indicates the relation between the aptitude and the response of the subject for each item. This response can be explained through a function that establishes the probabilities of a correct response called Item Characteristic Curve (ICC). Some of the parameters used for test questions are [6]:

- Difficulty. This factor describes the aptitude required for correct response of the item.
- Discrimination Level. This factor indicates up to how much the item permits the differentiation of subjects with a low aptitude and subjects with a high aptitude in responding to the item.
- Guess Factor. Represents the probability of a student of low aptitude correctly responding a question.
- Distraction Factor. Represents the probability of a high aptitude student incorrectly answering a question.

Question Bank. It is a fundamental element for the creation of a CAT. To define an efficient test banc we need to specify the various knowledge areas of the set. The test bank questions should contain questions of varying difficulty and variety [4].

Entry knowledge level. Proper selection of the first question so that the length of the test can be sensibly reduced [5].

Method for question selection. An adaptive test selects the following item as a function of the estimated knowledge level and the responses of the previous items by the student.

Termination criterion. Deciding when to end a test has several criteria like when a determined precision level has been achieved in relation to the knowledge level, and when a determined number of items have been set, etc.

2.2 Algorithm

A CAT is an iterative algorithm (Fig. 1) that starts with an initial knowledge estimation of the student and has the following steps [6]:

- All the questions still to come are examined so as to best determined which question will follow base don the student's estimated knowledge level.
- The question is selected and the student answers it.
- According th the response, a new estimation is performed of his knowledge level. Steps 1 to 3 are repeated until the termination criterion is met.

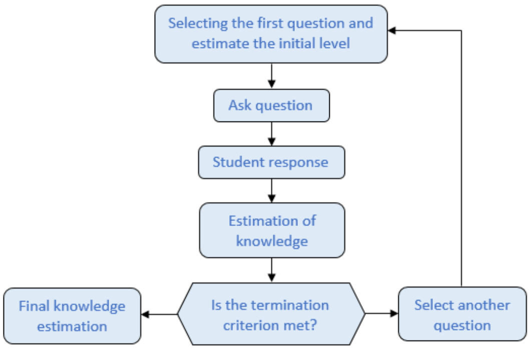

Fig. 1. Flow diagram of an adaptive test [1].

3 CAT Characteristics

3.1 Bayesian network structure

The Bayesian network used in this Project follows the specifications explained in [6]. In this sense, we'll have two types of Bayesian networks that are modeled with a different type of Bayesian network explained as follows:

a. Static Bayesian network

The evaluation process done on all the tests taken by the student permit us to know the level of knowledge obtained by the student during the diagnostic process. A classic Bayesian network (static) models this process. An example of this type of Bayesian network is presented in (Fig. 2). Next, the variables used in this Bayesian network are explained:

- Concepts: which relate to the evaluated vocabulary words. These are lower level items, for example: cat, dog, chair, board.
- Topics: a set of concepts that have something in common. A topic is equivalent to a test. These are middle level items, for example: Animals, School.
- Unit: A set of topics and correspond to the final evaluation results for the student. These items are higher level, for example: Basic Unit.

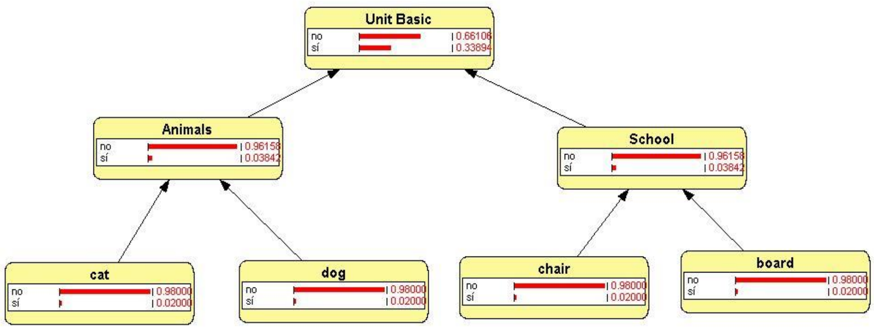

Fig. 2. Example of an evaluation Bayesian network (Basic Unit)

As explained in [7] we consider all the nodes to be binary and the probability interpretation of the different node type as follows: the concept nodes, represents the probability that the concept is known or unknown. In the assignment and topic nodes, the probability is interpreted as a measure of degree of knowledge obtained in the topic and assignment.

b. Dynamic Bayesian network

During a test, diagnostic process evidence is collected from the questions to estimate the level of knowledge of the test concepts. This process uses a dynamic Bayesian network since the evidence nodes change with time. It's not necessary true that if a student answers correctly a question related with certain concepts that he will always answer correctly these type of questions. [8].

The relationship between two successive dynamic Bayesian network levels are shown in (Figure 3). A continuous line represents the connections between the levels. A new type of bayesian network can emulate the behaviors of this dynamic Bayesian network. This new Bayesian network has the advantage that its use and implementation are simpler while its behavior is equivalent. In [6] we call refresh the process of refreshing the node is defined in the following manner:

Step 1. We refresh the probabilities of all nodes in the network given evidence E.

Step 2. We replace the previous node with a new node in which the probabilities a priori of the nodes with parent is equated with those calculated previously. Further, the conditional probabilities of the remaining nodes does not change. Thus, the same node where the evidence was acquired remains available.

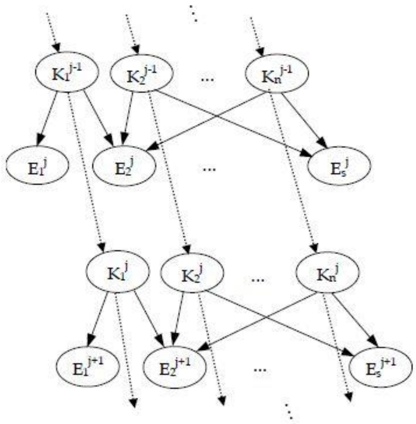

Fig. 3. Dynamic Bayesian network (knowledge and evidence nodes)

We can observe a simple example of this Bayesian network in (Fig. 4). We can explain the variables that participate in this type of Bayesian network.

- Concepts: Is the set of concepts of a particular topic, for example: cat, dog.
- Questions: Is the set of questions related to the topic. From these variables, we obtain the necessary evidence to calculate the degree of knowledge of the concepts in the example: Preg1, Preg2, and Preg3.

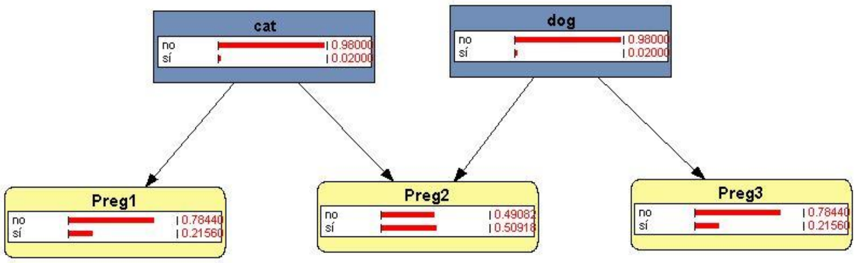

Fig. 4. Example of a diagnostic Bayesian network (Animals)

Furthermore, the relationships need to specify:

Probabilities a priori of the concept nodes. According to the initial level specified by the adaptive test.

Conditional probabilities of the topics, given the concepts, and the unit given the concepts. These probabilities can be obtained from a reduced set of parameters [6].

The set of values that measures the importance of each topic concept, and each unit topic. For this, let's suppose C_{ij} with $i=1,\ldots,n_j$ is the set of concepts related to the topic T_j, y w_{ij} represents the importance of concept C_{ij} in the topic T_j. Thus, the total probability distribution is calculated as shown below (Equation 1):

$$P\left(T_j \middle| \left(\{C_{ij}=1\}_{i\in S}, \{C_{ij}=0\}_{i\notin S}\right)\right) = \frac{\sum_{i\in S} w_{ij}}{\sum_{i=1}^{n} w_{ij}}. \quad \textbf{(1)}$$

Similarly, for each unit U given the set of related topics from such assignment Tj / $1 \le j \le s$, and for each $j = 1,\ldots,s$ given α_j the weight that measures the importance relative with the topic T_j in the unit U. Thus, the probability conditional distribution of U is calculated as shown below (Equation 2):

$$P(A|(\{T_i=1\}_{i\in S}, \{T_i=0\}_{i\notin S})) = \frac{\sum_{i\in S}\alpha_i}{\sum_{i=1}^{S}\alpha_i}. \quad \textbf{(2)}$$

Conditional probabilities for each question given the concepts shown within. These probabilities are divided in two parts, which are explained latter in the answer model of the computerized adaptive test.

3.2 Element details from CAT

We will now explain how we defined the elements of the computerized adaptive test.

a. Answer model

Many functions show this behavior. Currently we use the so called Logistical Models, based on the logistic distribution function. The most widely used is the Logistical Model with three parameters, in which the Item Characteristic Curve (ICC) of the ith question is described with a guessing factor, difficulty level, and discrimination factor.
For questions type multi-answer, the function is determined by the linear transformation of the function ICC [6], shown below (Equation 3)

$$G(x) = \frac{(1-c)(1+e^{-1.7ab})}{1+e^{1.7a(x-b)}} \quad \textbf{(3)}$$

.

Where:
c = Guess factor (l/n)
a = Difficulty level.
b = Discrimination factor.

G is calculated as described in (Equation 4).

$$\left\{G(0), G\left(\frac{x*}{p-1}\right), \dots, G\left(\frac{(p-2)x*}{p-1}\right), G(x*)\right\}. \quad \textbf{(4)}$$

Considering that G(x*) = 1 – s, where s is the distraction factor.

For the questions, the (Equation 5) determines type assignment the function explained in [6].

$$P(E=1|C_1, \dots, C_n) = \prod_{i \in S}(1-s_i) \prod_{i \notin S} g_i \, . \quad \textbf{(5)}$$

Where:
si = Distraction factor.
gi = Guessing factor.

However, we can simplify this formula since the used format for the question type assignment are related to only one concept. The final form is shown below in (Equation 6) and (Equation 7).

$$P(E=1|C=1) = 1-s \, . \quad \textbf{(6)}$$

$$P(E=1|C=0) = g \, . \quad \textbf{(7)}$$

b. Question Bank

For each question, the following parameters where defined: discrimination factor, difficulty level, and guessing factor. Additionally, the distraction factor was specified in the ICC calculation. So, this parameter was also established for each question.

The test is directed for children of 5 and 6 years old. For this reason, we are working with questions that do not have many items. Hence, the questions are not very complicated.

c. Initial Level

We used a uniform distribution for the initial level for all students. It's assume that no concepts are initially known.

d. Method for question selection

Millan [6] proposes several question selection criteria evaluated by simulated students. The author indicates that the best results were obtained with the conditional criteria to the question probability; we have selected it for our work.

The conditional criteria to the question probability is based in directing the diagnostic in the same behavior line shown by the student in the previous formulated questions where the utility is calculated by (Equation 8).

$$U(P,C) = \begin{cases} \text{máx}_{C \in pa(P)} P(P = 1/C = 1) & si\ P(P = 1) > P(P = 0) \\ \text{máx}_{C \in pa(P)} P(P = 0/C = 0) & en\ otro\ caso. \end{cases} \quad . \qquad \textbf{(8)}$$

e. Termination criteria

It was decided to use two criteria for test termination. The first criteria is applied when all the concepts have been evaluated. The second criteria is applied when the child elects to terminate the test. The purpose is to evaluate the student, as a learning method, to based on the answers proper tutoring is given. This way knowledge of the evaluated concepts increases.

4 Implementation

4.1 Implementation tools.

Sever: For testing a laptop computer was used with the help of software packages like Xampp, and Tomcat 7 as server applications.

Tools used for database management: For database management MySQL we utilized the tool PhpMyadmin.

Tools for the construction of Bayesian networks: For Bayesian network construction we used the tool OpenMarkov[1]. We used the library as an application program interface (API) for the creation, edition, elimination, and inference of Bayesian networks.

Tools for changing text into voice: To add voice functionality in the web system we used the library ResponsiveVoice[2]. This is a JavaScript library base don HTML5 that helps in adding voice functions in websites and applications. It functions in any device that has internet connection. It has no dependencies and is very light.

Tools for voice recognition: For the capture of voice, we used the tool Annyang[3]. We compared the chain captured by the tool with the chain of the correct answers and based on the comparison the results are shown, indicating whether there the response was correct or not.

4.2 System Architecture

In (Fig. 5) we can observe the general architecture of the developed web application, which was called Teaching and Evaluation System for the English Language (TESEL).

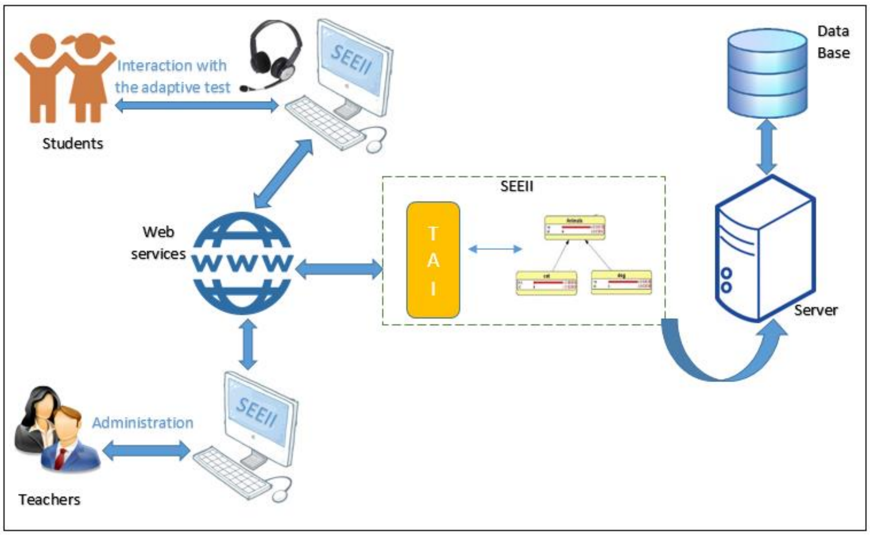

Fig. 5. General application of the TESEL application

For implementing the developed application, we worked as indicated in (Fig. 6), where we can visualize the purpose of each component of the system and how they communicate.

[1] OpenMarkov: http://www.openmarkov.org/

[2] ResponsiveVoice: http://responsivevoice.org/

[3] Annyang: https://www.talater.com/annyang/

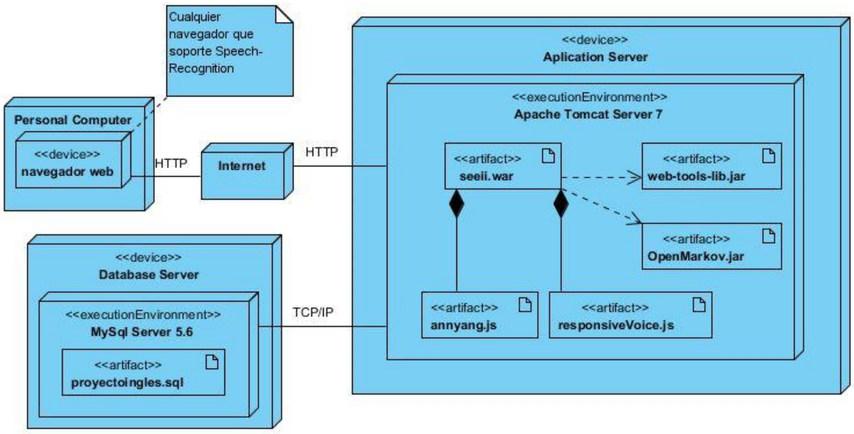

Fig. 6. Initiation diagram of the TESEL application.

4.3 Codification

Creation of Bayesian networks: Node administration and assignment of probabilities was done using the OpenMarkov Library. The code was developed following the specifications for coding Bayesian networks in the ProbModelXML format [9], and adapting them to Java which is used in OpenMarkov. The code used in the developed application can be found in: https://github.com/kiramonc/seeiiProjectF

5 Results

Fort the testing a simple of 20 students were used from the primary school Zoila Alvarado de Jaramillo. The children belonged to a first grade class. Results obtained are explained in details below:

5.1 Use of the web application by the students.

Work was done with one unit (Basic Unit), directed to children between 5 and 6 years of age. A topic was used for the evaluation (Farm Animals). This topic consisted of six concepts: horse, chicken, cow, sheep, pig, and rabbit. The interface is presented below.

In (Fig. 7) we show the screen with the available topics in the teaching unit that the student is registering.

Fig. 7. List of available topics.

In (Fig. 8) we can observe the screen presentation of the question that will evaluate the ability of the student to listen. A specific item is requested; in order for the program to read the question, we used the library ResponsiveVoice.

Fig. 8. Listening question evaluation.

In (Fig. 9) we can observe the presentation for the question that evaluates the ability of the student to pronounce. This question shows one or several items and the program asks the student to pronounce the name of the item. For voice capturing the library Annyang was used.

Fig. 9. Speaking evaluation.

(Fig. 10) shows the screen view for the results obtained by the student in the unit topics.

Fig. 10. Unit results.

To get a better picture of the obtained results with the web application we can observe in (Fig. 11) the medians of the obtained results by the students for each concept.

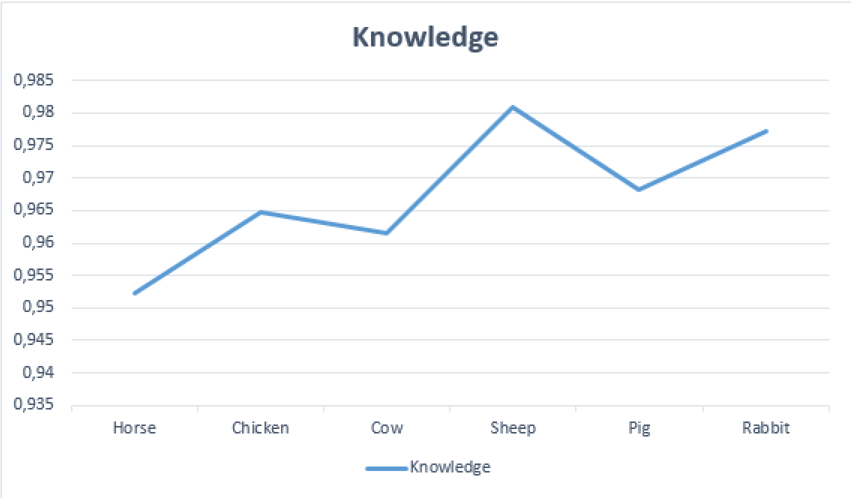

Fig. 11. Medians for each concept.

As we can see, the results are always greater than 0.95 (in the 0 to 1 interval). The worst result is for the concept Horse (0.9523), and Sheep has the best result (0.9808), which are satisfactory results. In (Fig. 12) we can see the new distribution of students within the three profiles based on obtained results.

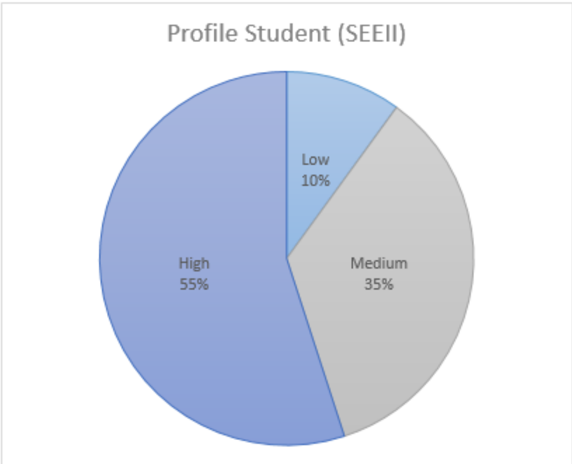

Fig. 12. Student Classification.

As we can observe the percentage of students with HIGH profile is 55%, which shows a great gain, whereas the percentage of students with a MEDIUM and LOW profile are 35% and 10% respectively.

6 Discussion

The present work shows that using Bayesian Networks as a method for student adaptation favorable results are obtained. The results are twofold, first we use a proper tutoring method and secondly we avoid frustration from students with little concept knowledge. Whereas, students with better concept knowledge find a higher degree of difficulty with more complicated questions and thus boredom is avoided.

Furthermore, the developed web application shows notable improvements for English evaluation since pronunciation is considered using voice recognition. This important characteristic makes our system a valid tool for estimating English language knowledge.

The developed application can be get better by adding more functionalities, so that more personalized test options in the administrative options. For example, we can add capacity for administering test formats, and inputting content that can be used for tutoring. Help menus can also be presented to the student that have low scores, these can be based on artificial intelligence techniques to keep the adaptability characteristic of the student.

7 Conclusions

- Based on information consulted it was determined that the conditional probability criteria of the question, is the best method for selecting questions to the level of student knowledge.
- Static Bayesian network allows diagnosis of the student as if a dynamic Bayesian network is used, but in a simpler way.
- The results obtained show that three levels of granularity (Unit, Topic, Concept) in the evaluation Bayesian network are enough for estimating the knowledge level of a student.
- With the use of voice recognition tool Annyang could evaluate the oral productive skills (Speaking) of the concepts defined in the computerized adaptive test.
- The elimination algorithm used for inference in the library OpenMarkov guarantees a fast inference, obtaining results in less than a second in a big Bayesian network.
- The web application TESEL helped students develop skills to listen and speak English.

Referencias

1. Sánchez J. P.: Agente selector de estrategias de enseñanza - aprendizaje para la educación basada en competencias. (2012)

2. López J., García J.: Sistemas de Tutorización Inteligente Basados en Redes Bayesianas. Universidad de Álmeria, (2008)
3. Conejo R., Millán E., Pérez J. y Trella M.: Modelado del alumno: un enfoque bayesiano, Inteligencia Artificial. Revista Iberoamericana de Inteligencia Artificial, vol. 5, nº 12, pp. 50 a 58, (2001).
4. Flaugher R.: Computerized Adaptative Testing, Hillsdale, NJ: Lawrence Erlbaum Associates Publishers, (1990)
5. Thissen D., Mislevy R.: Testing Algorithms, Hillsdale, NJ: Lawrence Erlbaum Associates Publishers, (1990)
6. Millán E: Sistema bayesiano para modelado del alumno, de Departamento de Lenguajes y Ciencias de la Computación, Univ. de Málaga. Málaga, (2000)
7. Millán E., Pérez J. y Suárez E.: An Adaptative Bayesian Network for Multilevel Student Modelling, de Proceedings of 3rd International Conference on Intelligent Tutoring Systems ITS 2000, (2000)
8. Reye J.: Two-phase updating of student models based on dynamic belief, Lecture Notes in Computer Science, nº 1452, pp. 6 a 15, (1998)
9. Arias M., Diez F. J. y Palacios M. P.: ProbModelXML. A format for encoding probabilistic graphical models, de technical report cisiad-11-02, UNED. Madrid, España, (2011)