An approach to knowledge assessment in an Intelligent Tutoring System

Constanza R. Huapaya
Francisco A. Lizarralde
Graciela M. Arona
Jorge R. Vivas

Basic Processes, Methodology and Education Research Center
Artificial Intelligence Research Group
Universidad Nacional de Mar del Plata
(7600) Mar del Plata. Argentina
huapaya@fi.mdp.edu.ar
flizarra@fi.mdp.edu.ar
grarona@fi.mdp.edu.ar
jvivas@mdp.edu.ar

Abstract

In this paper, we present an approach to student's evaluation in a well-defined domain based on a semantic network. A similarity matrix based on the semantic memory structure of humans is used to build a semantic distance model in order to describe an assessment technique to evaluate the student's state of knowledge. Our aim is to facilitate a deeper conceptual understanding of domain principles. We are developing a new student model including an assessment module with DistSem model.

Keywords:
Knowledge assessment, semantic distance model, Intelligent tutoring system

1 INTRODUCTION

The goal of ITS (Intelligent Tutoring System) is to provide the benefits of one-on-one instruction automatically and cost effectively. Like training simulations, ITS enables participants to practice their skills by carrying out tasks within highly interactive learning environments. However, ITS goes beyond training simulations by answering user questions and providing individualized guidance. Unlike other computer-based training technologies, ITS systems assess each learner's actions within these interactive environments and develop a model of their knowledge, skills, and expertise. Based on the learner model, ITSs tailor instructional strategies, in terms of both the content and style, and provide explanations, hints, examples, demonstrations, and practice problems as needed.

In order to provide guidance and instructional feedback to learners, ITS systems typically rely on three types of knowledge, organized into separate software modules:

- The *expert model* represents subject matter expertise and provides the ITS with knowledge of what it's teaching.
- The *student model* represents the student's knowledge, skills, and other attributes that affect how the student should be taught. This model lets the ITS know who it's teaching.
- The *instructor model* enables the ITS to know how to teach, by encoding instructional strategies used by the tutoring system.

The analysis of students' knowledge level poses a difficult problem mainly because knowledge is measured in terms of skill mastery, which is an unobservable abstraction. Some current approaches that overcome these problems are Bayesian networks and belief networks. Bayesian networks [1] evaluate an observable state of student behavior during problem solving. We propose a model based on semantic distance in order to acquire an approach of students memory map [2], specifically, measure the distance among nine concepts of the subject. Our aim is to reach a flexible and reliable assessment technique to include in student model of Intelligent Tutoring Systems.

Moreover, in Argentinian universities, the teachers do not have many opportunities to continuously follow the students' progress due to limited classroom time and crowded lecture rooms. With the proposed methodology, an ITS can assist the teacher in "closely" following each student's progress. We are currently developing an ITS in the first year of the Engineering course of studies, mainly the assessment module [3].

2 SEMANTIC NETWORK

A semantic network is a way of representing relationships between concepts. Often each concept is represented by a word or set of words. A simple example is a hierarchical network where the concepts are taxonomic terms from biology, and the only type of relationship is *is_a* or *type_of.* More complex semantic networks many include a variety of types of relationship such as hardness, temperature, texture and color.

In a semantic network two concepts are semantically related if they are close in a network. We can measure the "proximity" or "similarity" as the distance between the two concepts, this is the geodesic distance of the paths that join them together. This technique produces unsatisfactory results in the hierarchical semantic networks, since nodes that are hierarchically dependent, thus more distant, may be likely to result closer than nodes located in the same hierarchy. The

Spreading-activation Theory of semantic processing [4] is a search and comprehension reticular model in the human memory. The search is seen as triggering propagation from two or more conceptual nodes to their intersection. The priming effect is explained in terms of the triggering propagation from the first concept node and constitutes the basic process on which comprehension is settled.

When a person estimates the semantic similitude between two or more ideas he or she can establish different types of semantic relationships between them. Their proximity can be given because both concepts present an inferential logical relation, moreover both concepts can share many attributes by means of which relationships, not necessarily logical ones, can be established. The similarities in the shared attributes between two concepts can promote the establishment of analogical relationships that are facilitated by the presence of activation in their respective labels. Some studies about semantic relationships [5] [6] suggest that different cognitive processes allow to elicit part-everything, contrast, cause-intention, etc. relationships.

Computer implementations of semantic networks were mainly developed for artificial intelligence and machine translation, but earlier versions have long been used in psychology and linguistics.

3 A SEMANTIC DISTANCE MODEL: DISTSEM

The Distsem method [7] consist of the use of the Social Analysis Networks [8] where the canonical use of the nodes as social agents is substituted by bonds such as semantic distance between the nodes-concepts, thus obtaining a semantic network. This method captures the semantic proximity estimations given by the participants and their inclusion in a matrix, on which it is possible to visualize and compare qualitatively and quantitatively the paticipants' semantic networks with only one level of restriction given by the limited number of previously defined concepts.

Specifically, Distsem allows to extract the constitution of a semantic network based on the distances between meanings, constitute their semantic matrix, describe, analyze and visualize their relationship and distribution in two dimensions and compare different matrices among each other and evaluate their proximity with a configuration proposed as correct.

The method is developed from stages where the following processes are carried out:

Stage 1: matrices and forms preparation. The n concepts are selected whose semantic entailment is desired to know. A squared matrix of concepts against concepts is generated. According to the nature of the problem, the best scaling instruction of pairs of concepts according to their similarity /dissimilarity and the Administration Form is prepared. Each pair of concepts resulting from the crossing of concepts against concepts is put in a form. The amount of pairs results from applying the following formula: $(n * (n-1)) / 2$ for non-directional relationships. Four repeated pairs with inverted order are added to evaluate the internal consistency. The pairs are ordered randomly for their presentation.

Stage 2: ddministration. Experts and students are asked to estimate the similarity (proximity) among the pairs of concepts presented in the Administration Form.

Stage 3: evaluation. As the resulting matrix of each concept is defined by a vector constituted by the values in respect to the other n concepts, estimated by each subject, one obtains the geodesic

distances matrix and a multidimensional metric scaling of objects is applied, in this case concepts, to generate a bidimensional semantic space for each participant. In order to know the produced semantic sets, the Hierarchical Cluster Analysis is applied to each matrix based on Johnson's process [9]. In this way, the groupings with the most cohesion (less distance) are generated among subgroups and their relationship with the totality.
In order to quantitatively compare the similarity among matrices produced by the participants among each other or against the expert's matrix, the QAP method (Quadratic Assignment Procedure) is applied.

Stage 4: results analysis. The described procedure allows for different perspectives and different level of analysis according to the researcher's interests:

- Visualize the semantic network that bonds the concepts. Appreciate the strength of their connections according to color and the outline of the bonds. The relationships that constitute associated ideas are grouped in nodes with the same color.
- Qualitative evaluation: Absences, excesses and impertinences of links among concepts can be visualized.
- Quantitative evaluation: It allows to measure the level of similarity among each participant's semantic network in different moments, their group and with a matrix considered to be correct.

The results of such treatments give rise to the description, analysis and visualization of the semantic network. Furthermore, DistSem gives rise to the comparison between two semantic networks, considering one of them to be the pattern. The DistSem method for the evaluation of semantic distances constitutes a broad and flexible procedure that allows for different levels of analysis as much qualitative as quantitative.

4 TESTS WITH DISTSEM

The implementation of different evaluation means such as multiple choice, true/false, etc. involve the programming of specific components, as well as the acquisition and processing of the information produced by the student. The ITS have posses algorithms to analyze these results and decide the following action the student has to execute.

This type of evaluation mechanisms are the most used due to its implementation and processing simplicity. However, they present some limitations based on the insufficient representation of the cognitive activity.

The evaluation of a first year Engineering student's knowledge begins with a new test based on DistSem. The teacher proposes nine main concepts related to the topic to be evaluated. For instance, in Figure 1a we can see part of a table on topics of the subject Numerical Analysis course. The student has to estimate how related the 40 pairs of concepts are, as we see in Figure 1b.

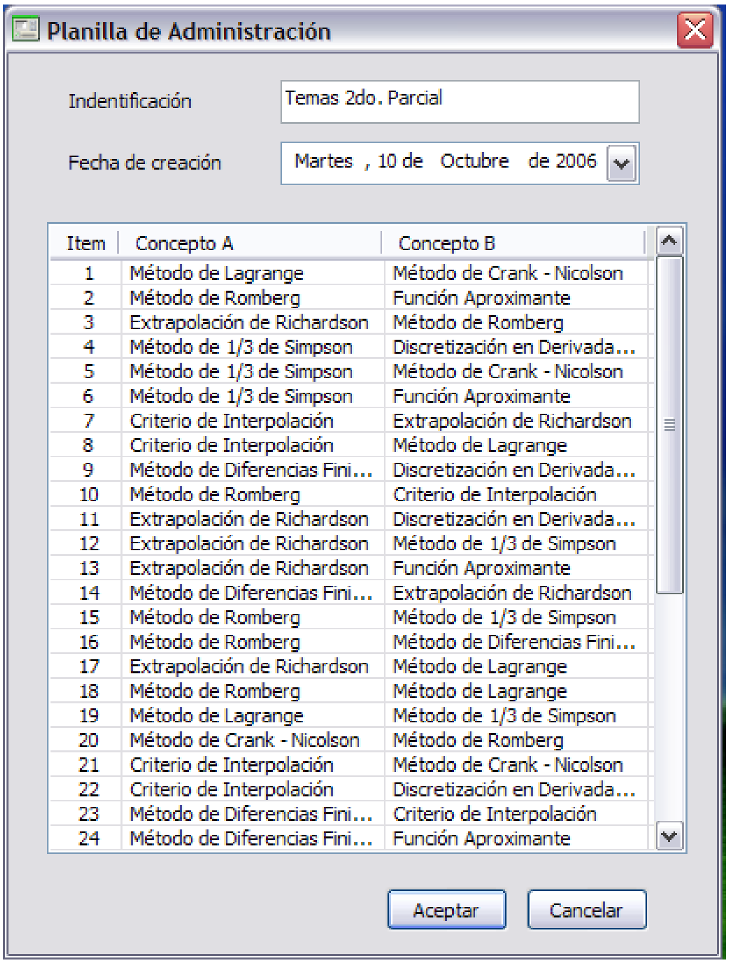

Figure 1a. DistSem test: administration table

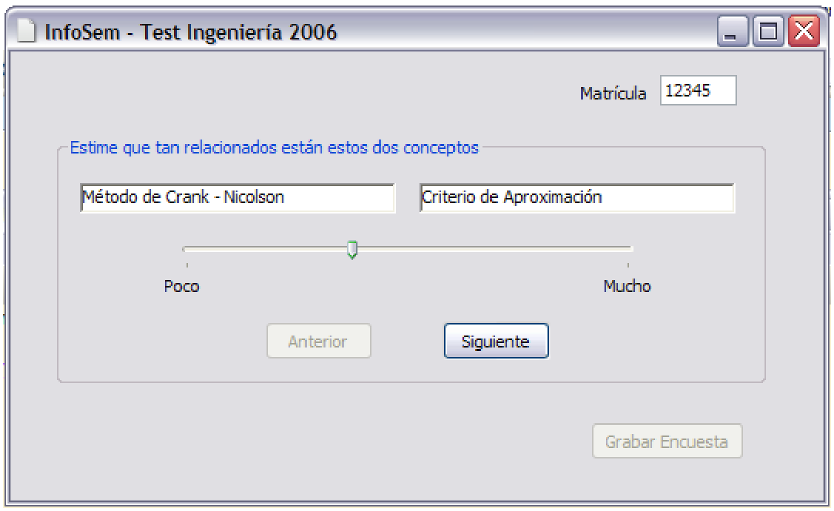

Figure 1b: student choice

From the student responses, a similarities symmetric matrix is built which reflects the student's semantic network (Table 1). Similarly, a pattern semantic network proposed by the instructor/expert can be built. The comparison between the networks of the expert and that of a particular student will take part in the evidence that will guide the ITS to decide the mark of its Global Instructional Model.

Table 1: similarities matrix

	Lagrange	Crank-Nicolson	Romberg	....	Interpolacion
Lagrange	7	6	2	....	4
Crank-Nicolson	6	7	1	………	1
Romberg	2	1	7	....	3
Interpolacion	4	1	3	….……..	7

The adaptation of the DistSem method to our ITS for the evaluation of the student's knowledge captures, apart from the comparison of the expert's and the student's understanding, the significance of the association between matrices created by each of them through the use of QAP. Considering that the data are dyadic and interdependent (40 pairs of concepts), we apply the Quadratic Assignment Procedure (QAP) [10] aiming at measuring the significance of the observed correlation between the expert's and the student's matrices.

QAP is a nonparametric permutation-based procedure, that preserves the interdependence among dyads. The pure procedure should generate 362880 (9!) permutations on the students' matrix and the expert's matrix should remain the same. Aiming at improving calculation times, the system modifies the original procedure randomly selecting the calculation from 2500 permutations. Afterwards, a correlation is found. We look for the level of significance by using the corresponding correlation to the student 's matrix in the found distribution. The percentage of the number of permutations corresponding to that correlation gives us the level of significance ρ. Afterwards, a correspondence with our Argentinian traditional grades between 0 and 10 is built.

The evaluation process of the student's knowledge about certain topics can be facilitated by using a criterion associated to the objectives and competences that establish what aspects of the topic must be known by the student. Following the criterion evaluation based on instructional / competence objectives, we have used the proposed by the teacher in order to reason on the learning levels reached by the students.

In our ITS several standard or Distsem tests about specific topics can be defined, and we have considered three instructional objectives [11] (Bloom 1956): *knowledge*, *comprehension and application*. For example, the topic “Linear convergence of a numerical method” can be evaluated from the point of view of one of these objectives.

Particularly, we have chosen the *comprehension* and *application* for the treatment of the student's diagnosis. The comprehension implies showing an understanding of terms and concepts, i.e., use the acquired knowledge to understand new information. Knowing that comprehension is key for the acquisition of new knowledge, reaching that objective is of great importance for the engineering student in his learning. Among the actions the student does: explaining, describing, interpreting, classifying, discussing, generalizing, telling, selecting, summarizing, we add “relating concepts” according to the semantic distance as the main action proposed by the Distsem method.

The application of principles to concrete situations is considered a higher-order task. The assessment of this type of tasks requires a significant faculty involvement, namely, teacher's intuition, judgment and experience.

The confidence is implemented through the *confidence level* that evaluators use when entering their judgments into the system. If an evaluator is highly convinced of the accuracy of his or her judgment, he or she will enter a high confidence level. Otherwise, if an evaluator is very unsure of his or her judgement, he or she will enter a low confidence level. The complete range of this confidence levels varies between 0 and 100.

5 STUDENT MODEL IN OUR ITS

In our student module (figure 2) the traditional components are appreciated and the figures in grey are the components we have dealt with in this project. The students' profile, the individual session organizer and the tests repository, problems and exercises are not treated here.

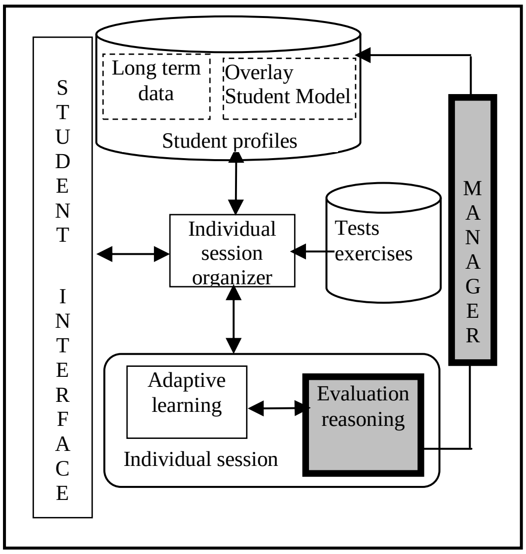

Figure 2: student module

The two components with the highlighted edges are the “Evaluation reasoning” where the DistSem algorithm is embedded and the “Manager” where each student “judgement record” is kept.

The Distsem administration forms as well as the expert's forms are stored in the data base. Moreover, traditional tests to be evaluated by the teacher can be included.

A modification of the Distsem algorithm is stored in the “Evaluation reasoning” component, that is ti say:

1. Expert's distance estimations matrix construction based on his or her form.

2. Student's data gathering with the administration form.
3. Student X's distance estimations matrix construction based on his or her form.
4. Exchanges produced on student X's matrix calculation using the QAP method.
5. Pearson correlation coefficient calculation in order to measure the significance of the association between the expert's and the student's knowledge.

Pearson's coefficient is a real value that varies between -1 and 1. The correspondence used in the system takes the positive values assimilating the subintervals to a whole value following the following rule: (0,0.1) to 1; [0.1,0.2) to 2; ……..; [0.9,1.0] to 10. In this way, a judgment score inferred by the ITS can be obtained. Furthermore, the system allows the evaluator to propose a confidence level in the Distsem test.

In the case of traditional tests, the human evaluator indicates a number from 0 to 10 to grade the test and selects a confidence level.

In all cases, the confidence level is used to assist in the inference processing.

If we assume that the evaluator will grade N tests, either Distsem or traditional ones, with confidence judgments for each one, the following normalized formula is calculated:

$$\text{judgment weight}_i = \frac{\text{judgment confidence}_i}{\sum_{k=1}^{N} \text{judgment confidence}_k}$$

Finally, the inferred judgment score is determined using a weighted sum:

$$\text{Final inferred judgment score} = \sum_{i=1}^{N} (\text{judgment score}_i \quad \text{judgment weight}_i)$$

The “Manager” component contains a “judgment record” where several fields for representing an evaluation are kept. Each student has the the following data stored for a set group of topics that are tested with Distsem tests or traditional tests:

- Student's Identifier and evaluation date
- Evaluated topics (that are part of the semantic network)
- Area (formative or summative)
- Type of evaluation (exam, preliminary exam, etc.) (they are N evaluations)
- Similarities matrix
- Judgment score of evaluation (Significance of the association between the student's and the expert's semantic networks or teacher’s grade)
- Level of confidence in the teacher's judgment (per evaluation)
- Teacher's commentaries (text) on his judgment <optional>
- Evaluation final inference given by a weighted sum

In order to support decision-making with rules of production, the ITS infers the student's level of knowledge about the group of topics. In our case, the comprehension level is categorized as *low, regular, good, very good.* This category affects the system's adaptive tutorial strategies, especially the feedback to be presented to the student, the following topic the system has to teach and the hints to use. The user can appreciate the expert's and the student's semantic networks in order to qualitatively compare the two, as is shown in figures 3 and 4 (“relating concepts” according to the semantic distance). The Pearson coefficient in figure 4 is 0.45.

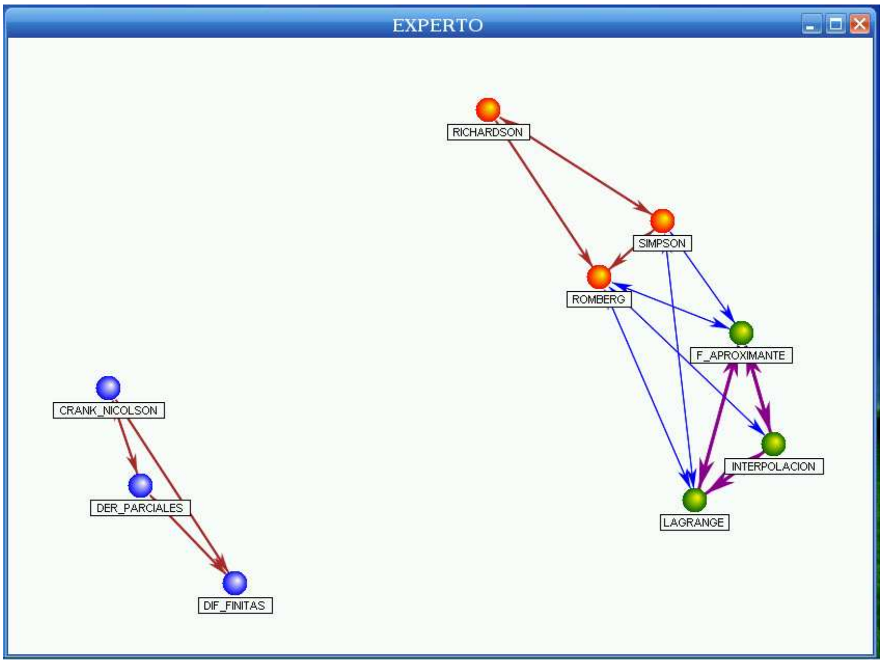

Figure 3: expert semantic network

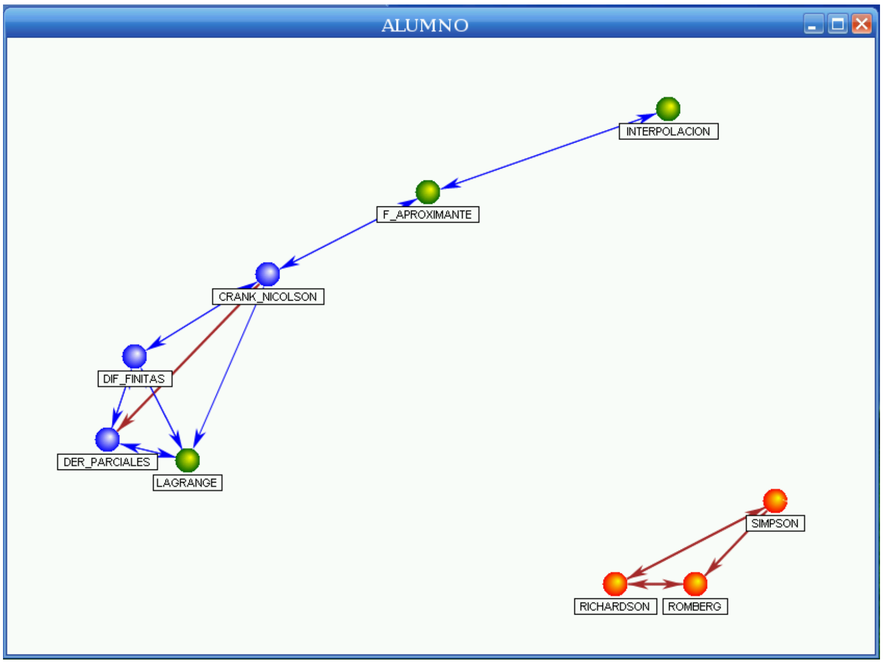

Figure 4: student semantic network

Furthermore, the rules of production give back the final results of the evaluation to the individual overlay student model where each node corresponding to the topic is marked as partially, totally or not learnt.

After each student/ITS session, the stored DistSem tests are updated according to the test's success or failure during its performance. This procedure improves the knowledge of the system about student's progress, and thus the tutorial strategies, reach a better effectiveness level (in the pedagogical module).

5 CONCLUSION

The new design of our Intelligent Tutorial System involves a student model containing the DistSem method. The inclusion of this method improves the ITS teaching effectiveness. It mainly enhances the quantitative and qualitative evaluation of knowledge with a technique based in a semantic distance model between concepts. The pursued objective is to reach the evaluation of comprehension and the skills of thought at a high level, as aplication. Furthermore, the DistSem technique contributes to the system' decision making in order for them to understand, classify and give feedback to the students.

6 REFERENCES

[1] Conati,C., Gertner, A. and VanLehn,K . Using Bayesian Networks to Manage Uncertainty in Student Modeling. *Journal of User Modeling and User-Adapted Interactio*n, 12:371-417. 2002.

[2] Quillian, M. Semantic memory. En M. Minsky (Ed.), Semantic information processing, Cambridge, MA: MIT Press. 1968.

[3] Huapaya C. and Arona G. Sistema de autoría especializado en STIs matemáticos, *Revista Brasileira de Informática na Educaçâo,* 10(2):37-49. 2002.

[4] Collins A. and Loftus E. A spreading-activation theory of semantic processing, *Psychological Review*, 82:407-428. 1975.

[5] Bejar, I.I., Chaffin, R. and Embretson, S. A taxonomy of semantic relations. In I.I. Bejar, R. Caffin y S. Embretson (Eds.) *Cognitive and psychometric analysis of analogical problem solving.* . 56-91. New York: Springer-Verlag. 1991.

[6] Mayor, R. and López, R. Relaciones Semánticas. In *Anexos de la Revista de Psicología del Lenguaje. Anexo 2.* Madrid: Departamento de Psicología Básica. U.C.M. 1995.

[7] Vivas J. Método Distsem: procedimiento para la evaluación de distancias semánticas. *Revista Perspectivas en Psicología,* 1:56-61. 2004.

[8] Wasserman S. And Faust K. Social Network Analysis. Methods and Applications. Cambridge University Press. 1998.

[9] Johnson, S. C. Hierarchical Clustering Schemes. *Psychometrika*, 2:241-254. 1967.

[10] Krackhardt D. *QAP partialling as a test of spuriousness*, Social Networks 9: 171-186. 1987.

[11] B. Bloom, Taxonomy of Educational Objetives, The Classification of Educational Goals: Handbook I, Cognitive Domain, Longmans, 1956.