The Design of an Intelligent Tutoring System Using the Natural Language Processing Technologies

Iuliana Dobre

(1) Petroleum-Gas University of Ploiesti
39, Bucuresti Blvd, 100680, Ploiesti, Prahova, Romania
E-mail: iulianadobre@yahoo.com

Abstract

Today’s society has increasingly more in need of learning techniques of high quality and, therefore, it has become increasingly clear the necessity to supplement the knowledge acquired by traditional means, taught by a teacher and using learning programs computer-assisted. The researchers performed in the Intelligent Tutoring Systems (ITS) represent a lever between artificial intelligence, cognitive science and education.

The main advantages of computer-assisted versus traditional tutoring are those that a computer never turn tired, the system is always available at any time and can be accessed by users from everywhere. In addition, compared with a classic computer-assisted tutoring system the ITS offers a personalised assistance and a real time feedback to the student, moulding itself on the student needs, and having as main purpose to enhance student’s learning grade, such target being easily achieved by building the model of the individual.

The purpose of this paper is to present an ITS designed by using Natural Language Processing technologies. The ITS main scope is to assist the users in their learning process for the Computer Programming discipline.

Keywords: Computer-aided instruction, instructional system, Intelligent Tutoring System, Natural Language Processing, personalized instruction

Introduction

As it was presented by the author of this paper in the first report issued during the PhD program, the Information and Communication Technologies (ICT) have recorded an extraordinary progress in the past ten years. Such progress has lead to a real revolution in the computer-aided instruction field (Dobre, 2010). “Amid the rapid changes and the technological progress achieved, as well as the trend to globalization of the higher education and the disappearance of the boundaries between students, new perspectives have been opened for the instructional practice. Thus, the instructional practice has been fitted with new methods, tools, applications and modern teaching-learning-assessment technologies, which can be considered today specific to informational society.” (Dobre, 2010).

The main advantages of the computer-aided instruction versus traditional instruction are many, the most important ones being the assurance of the instructional process continuity and of the universal access to instruction resources. Additionally, if a classic computer-aided instructional system is compared with an ITS, can easily be noticed that the ITS offers to student the option of a personalised instruction and of an immediate feedback about the performances achieved. Such options are available within an ITS due to the ITS capacity to adapt to the student needs being focused to support the increase of the student knowledge. The ITS adapting feature is made possible based on the modelling of the profiles of the individuals (students) who are using the ITS.
The scope of this paper is to present an ITS developed by the author of this article, using the Natural Language Processing (NLP) technologies.

**What Is an ITS**
As per Nwana definition the „Intelligent tutoring systems (ITSs) are computer programs that are designed to incorporate techniques from the AI community in order to provide tutors which know what they teach, who they teach and how to teach it” (Nwana, 1990).

In other words, an ITS is a package of programs capable to generate automatically and in an intelligent manner new instructional materials which will be introduced to student in conformity to his individual profile, needs and level of knowledge. Using techniques for automated instruction, the ITS is self-monitoring, self-assessing and self-improving its performances. Through the ITS any instructional activity could be initiated by any of the two, the student or by the system itself.

To develop an efficient instructional system it’s necessary to use methods, technologies and tools from three different primary domains as follows: Computer Science, Psychology and Education, this being showed in figure 1 (Woolf, 2009).

![Figure 1. The development of an ITS using methods and instruments from three different domains (as presented in Woolf, 2009)](image_url)

Each of the three domains have sub-domains which are contributing to the ITS development. For example, the Artificial Intelligence which, according to Woolf is a sub-domain of the Computer Science, contribute to the development of the intelligent features of the ITS. The Cognitive Science which is a sub-domain of the Psychology is contributing to the development of the means which will be used by students to process and assimilate the information received through the ITS. Meanwhile, the Education is used to find and to apply in the most efficiently way the most evolved methods for teaching-learning-assessment. The Woolf’s three primary domains common intersection resulted from their interfingering (see figure 1) is considered by Woolf the birth place of the ITS (Woolf, 2009).

**The ITS Structure**
The ITSs paradigm is closely connected to the huge progress recorded by the knowledge-based systems, called also expert systems. Starting from expert systems the specialists in instructional systems were looking to develop a new system, fitted with an explicit knowledge base, and which could be used in a different context than the one for which was built. This different context targeted by specialists was the instruction of the novices. The specialists considered to add some
specialised modules to an expert system so would be possible to transfer knowledge (and implicitly competence) from an expert to a novices. In a such case the system could be considered an intelligent system and the system will get smart instructional meanings (Ştefănescu, 2006).

This vision led to the building of a architectural structure which is typical to ITS. Such structure is consisting from four macro-environments as it’s presented in figure 2 (Nkambou et al., 2010).

According to researchers (Nkambou et al., 2010; Trăuşan-Matu et al., 2005; Nwana, 1990; Self, 1988), the four macro-environments are in fact models described as follows:

- **domain model** – contains concepts, definitions, rules from the learning domain, as well as strategies to approach the solving of the exercises and problems, base on which will be performed the student performances assessment, will be identified all mistakes and will be generated a personalised feedback. The knowledge-base of the learning domain could be represented in several modes with the support of ontology, hierarchies, semantic networks, frames, production rules (Trăuşan-Matu et al., 2005);

- **student model** – is containing aspects about student as follows: the initial knowledge and the knowledge acquired during the instructional process, the student’s personal data, the student’s performance evolution during the instructional process, information about student emotional and affective state. In some situations, the student’s knowledge base is included in the domain’s knowledge base. The importance of the student model in the ITS architecture has been highlighted by Self in 1988, when Self has defined the six major roles played by student model these being: corrective, elaborative, strategic, diagnostic, predictive, evaluative (Self, 1988 cited in Nwana, 1990);

- **tutoring model** – is the model that contains the learning and teaching strategies, examples and analogies. Based on the information provided by the student and domain models, the tutoring model choose an action, a strategy, a pedagogical technique or a combination of actions, strategies or techniques which could help the student to assimilate the maximum of information & data;

- **communication model** – is in fact the system interface with which the student could communicate with the computer.

In order that an ITS to be effective, the ITS has to be fitted with a very extended knowledge base of the domain targeted, and the pedagogical strategies used to assist the student during the instructional process must be more varied and more personalized so the needs of the student will be fully covered. It also has to make possible a permanent connection between student and tutor (teacher) in a manner that this connection will become a mean through they can communicate.

**Description of the Proposed ITS Architecture**

The proposed ITS architecture is based on the one proposed by Nkambou et al. and shown in figure 2 (Nkambou et al., 2010). The author of the present article has looking more to the practical side of the problem and keeping same ITS architecture has focused on the materialisation of the ITS using tools available on the Web and having public free licenses. More over, the author has chosen as domain of applicability the *Computers Programming and C Language* discipline.
The idea was to develop an ITS capable to assist through the entire instructional cycle (teaching – learning – assessment) those students who want to take onboard and to understand notions about Computers Programming and C Language. Another idea followed when the proposed ITS was developed was that the system should be capable to follow up the student evolution based on the results obtained during the assessment phases. Briefly, the assessment part was developed based on tests with questions to which the students have to answer in text format. The ITS is analysing the answers provided by the students using the NLP technologies and generates a personalised feedback to each of them.

The ITS proposed was developed using Java and MySQL together with the Web service for text processing made available by the Artificial Intelligence Institute (AII) from Bucharest (http://www.racai.ro/webservices/TextProcessing.aspx).

The Domain Model
The proposed domain model contains the knowledge base specific to Computers Programming course as it is taught in the Oil-Gas University of Ploiesti. Here have been stored all the information pertaining to this course in order to provide the student with the required knowledge.

The knowledge base from the domain model and specific to Computers Programming course covers to pics as follows: the evolution of the computing systems, the phases of a program execution, numeration bases, solving problems assisted by computer, elements, instructions and other aspects related to C Language.

The information stored in the domain model will be used later, during assessment phase, to generate the tests questions and to perform the analysis of the answers provided by the students, as based on this analysis will be calculated the scores and a personalised feedback will be generate and sent to students at the end of each assessment session.

The Student Model
The second module, the student model, identifies and memorise the information about the student: personal details, the student evolution up to date, the taken tests and the scores obtained. All information from the student model are updated permanently based on the progress recorded and all the information available at one moment are used to create the personalized feedback sent to each student. Also, the information stored start from the time when the student register within the ITS using the registration interface.

The proposed student model stores the following categories of data:

- **Student’s personal data & registration** – figure 3 illustrates which personal data are stored in the student model. These data are collected and stored from the first login done by the student. A graphic interface is used to collect these data and also, the interface is fitted with an option allowing the student to carry on any modification deemed necessary to be done. The data submitted are protected from any non-authorized access through a password given by student. As traditional for online access systems, the user name and the password are required by the ITS to recognize the student. Additionally, the system will require to choose between user and administrator level as the ITS access graphic interface it is used also by the system administrator. The student e-mail address is another mandatory information to be provided as the student e-mail address will be used to exchange important data and information (i.e., user registration confirmation, informing messages about aspects/topics/issues etc.). Figure 4 is showing a test example for user registration confirmation;

- **Student’s results after taken the tests** – the students are assessed by the ITS after each course stored in the domain model is completed. The scores obtained by the students are recorded in the student account created within the student model;
• **Student’s evolution history** – the student model is storing valuable information about student evolution during the instructional cycle such as: number of courses completed, number of tests taken, the scores obtained etc. The data are recorded after each course is completed and the attached test is taken by the student. Also, the ITS has the option to create graphs for a better understanding of the student’s evolution such as the graph of the results obtained calculated between two dates as requested by the solicitor. Also, overview (general) graphs can be created to have the picture of the student’s evolution at macro level at a desired moment.

**Tutoring Model**

The proposed tutoring model has been developed in a manner consider by the author of this article as an attractive one focusing also to have set up a systematic organization and having as final target to connect the student as user of the ITS with the instructional process respecting in the same time the student needs.

The tutoring model stores the following categories of data:

• **Course materials** – the students have at disposition all courses contained by the Computers Programming and C Language curriculum as it is in used at Oil-Gas University from Ploiesti. The graphic interface is easy to use and as it is shown in figure 5 offers to the student the option to start with the very first chapter by just selecting that one from the list but also offers the possibility to come back to any of the previous chapters as student deemed necessary;

Figure 3. Screen capture of the graphic interface – data page showing one fictive example of registration

Figure 4. Screen capture of the graphic interface – data page showing one fictive example of registration
• **Tests generation** – after each chapter the tutoring model is generating a test with ten questions from the chapter completed by the student. There is set up a time limitation for student to answer. The time allocated as well as the number and the content/difficulty grade of the questions could be modified by the system administrator as necessary. The questions are generated randomly;

• **Student answers analysis** – the students answers are analysed by using the AII Web text processing service (http://www.racai.ro/webservices/TextProcessing.aspx). The service compare the students answers versus the correct answers stored by the teacher in the system. The comparison of the answers is done using the BLEU algorithm (Papineni K. et al., 2001). The ITS has set up a minimum percentage of correct answers which shall be provided otherwise the student will not pass the test. There is the option to repeat the cycle and re-take the test at the end;

• **Tests summary** – the personalized feedback is provided either through graphs either through recap of the taken tests. The feedback is constructive and is continuously provided to both, student and teacher.

**Communication Model**

The communication model is in fact the ITS interface and was developed to be as much as possible attractive and systemically organized in order to be user friendly.

**Conclusions**

The author of this paper considers that the future of the instructional process belongs to ICT and to ITS. There is continuously increase in demands of hardware and software implementation in all organizations involved in instructional process. Higher education is not making any exception from this mega trend and therefore, all higher education organizations shall be prepared to sustain the development and the implementation of the ICT and ITS within their curriculum regardless the specializations and the disciplines.

In this paper the author has briefly presented a proposal of an ITS built in respect of the ITS classic architecture as presented by Nkambou et al. (Nkambou et al., 2010). The proposed ITS is using the text processing service provided by AII which is based on NLP technologies and on the BLEU algorithm. The author is still working to finalise all details regarding the proposed ITS and to have it ready for the testing phase, which is the next step which will be taken in the short future.

**References**

**Books:**


**Book Chapters:**


Journal Articles:

Technical Reports:

Theses:

Internet Sources: