Design and development of the curriculum for master of sciences in teacher of informatics

Ana (Resulaj) KTONA[1], Denada (Xhaja)[2] ÇOLLAKU, Alda KIKA[3]

Abstract

Teaching and learning of the subject of Informatics in pre-university education is as important as teaching and learning of other scientific subjects, but taking into account the dynamics of change and the broad areas of its implementation, it requires different methodologies and special attention. The completion of the knowledge so far, for pedagogical staff of Informatics, was conducted by sporadic individual or organized trainings. A master of science for teacher in Informatics, which provide to the teachers of this discipline the necessary knowledge and skills for the development of this subject in different cycles of the study, does not currently exist in our country.

The department of Informatics set up working groups for the identification of problems and designing the curricula that will be developed in this master degree program. This process was very difficult because there is no study which specifies in measurable and comparable way the level of the knowledge that students currently receive in high school and problems encountered during the course of Informatics. The department of Informatics therefore undertook this study in order to create the appropriate curriculum for the master of science "Teacher in Informatics for high schools".

Data were collected through a web site that serves for electronic data collection and through printed surveys. The most appropriate data mining techniques for this kind of data was determined after conducted a theoretical and experimental comparative study between various data mining techniques. The problems encountered during the course of Informatics were identified using this technique. Addressing the identified problems led to the curriculum design for Master of Science in teacher of informatics which will be organized at the department of Informatics, Faculty of Natural Sciences, Tirana.

The method and the measurable resulting index will serve as a starting point for studying the data that will continue to be collected in the coming years. The issues that will be identified and addressed in the future will lead to the curriculum development for

---

1. *Ana (Resulaj) Ktona, PhD - Faculty of Natural Sciences, University of Tirana.*
2. *Denada (Xhaja) Çollaku, PhD candidate - Faculty of Natural Sciences, University of Tirana.*
3. *Alda Kika, PhD - Faculty of Natural Sciences, University of Tirana.*

Master of Science "Teacher in Informatics."
Key terms: Informatics teaching, Informatics learning, data mining

Introduction

Informatics is a core subject in our secondary education but it is still a new subject. Its program has changed frequently in recent years. [1][2][3] The subject that provides to the students the possibility to gain new knowledge in the field of Informatics in high school or technical secondary schools is in the technology field and has the name of information and communication technology (ICT). [1] According to the new curriculum guide for the high school, ICT is a subject, but is more an interdisciplinary subject. [4]

Referring to the curriculum of ICT what is taught in the subject is more the use of ICT as a tool to help in other areas of life. Informatics field is constantly changing and it has applications in all areas of life. Therefore the knowledge of teachers and students should not be limited in the use of ICT as a tool. They should have the basic concepts of Informatics in order to cope with the constant changes in the field.

So far the knowledge completion for pedagogical staff of Informatics was conducted by individual or organized sporadic training. Currently it does not exist in our country a master of science for teachers in Informatics, which provide the teachers of this discipline with the necessary knowledge and skills for the development of this subject in different cycles of the study.

The Department of Informatics set up working groups for the identification of problems and designing the curricula which will be developed in this master. This process was very difficult because there is no study which specify in measurable and comparable way the level of the knowledge that students currently receive in high school and problems encountered during the course of Informatics. Therefore, the department of informatics undertook this study in order to create the appropriate curriculum for Master of Science "Teacher in Informatics for high schools". This Master will be organized at the Department of Informatics, Faculty of Natural Sciences, Tirana.

Methodology

In view of the main aim of the study we designed a questionnaire, which was tested in a two-week period to see the possible defects and to improve it. The questionnaire was available online after being published on a website that served for the electronic data collection. Data was also collected through interviews with high school students and students The interviews was performed with the purpose to complete the questionnaire. Through the questionnaire we collect data about geographic area, gender, laboratory base, etc, the level of knowledge they have about the different lines of the curriculum, student's motivation related to these lines, as well as their opinion about the importance of the subject and the received knowledge in the subject of Informatics.

In order to specify in a measurable and comparable way the level of knowledge that students currently receive at high school and the problems encountered during Informatics course we used the data mining techniques. Through data mining can be inferred the meaning of data: it can generate models from data called "training data" and apply them in other data.

We conducted a theoretical and experimental comparative study between the different methods and techniques of data mining to determine the most appropriate

technique for this kind of data. Among the studied methods are: classification methods, cluster methods and association methods.

Results and discussions

Initially we conducted a theoretical study between methods of data mining: forecasting, collection and report detection.

Classification methods are used when we want to build a model which gives us knowledge of a feature of data using a combination of other data features. Through these methods can be identified which data features are important to gain knowledge on a particular feature of these data. Therefore, to identify the features that affect more on the level of the gained knowledge we will use one of the forecasting methods. Through forecasting methods we can also calculate indices and measurable values to assess: the level of knowledge, the level of teaching, or the level of subjective satisfaction of the students.

Cluster methods are usually used when some data features naturally group together dividing so the set of data into some data sets. These methods are usually used when these groups are not known in advance and are not visible. These methods can be used to detect possible groupings in the collection data as a whole, which can be used to formulate personalized teaching methods.

Report detection methods are used to detect reports between data features in a data set with a large number of features. These methods can be used to detect the connection and the dependencies between data features.

Given our goal we will study the forecasting methods to discover the most appropriate method to understand the level of the acquired knowledge by high school students in Informatics and the data features that affect more in this level of knowledge.

Some of the methods that can be used to forecast are artificial neural networks, bayesian networks, decision trees.

Artificial neural networks are able to discover model and tendencies at data that can not be noticed by people or other computer techniques. In data mining the artificial neural networks can be used in problems of classification, forecasting and in groupings. The models that are used more from artificial neural networks in data mining are: feed-forward back propagation that is used for making classifications and predictions, and the Kohonen's self-organizing maps that are used for cluster analysis.

An artificial neural network is a system for processing information based on the biologic neural system. It consists of many simple processing elements (artificial neurons), which are organized in groups called layers. Every neuron is directly connected to other neurons and every connection is associated with a parameter called weight. Weight represents the knowledge acquired during the learning process. There are several models of artificial neural networks. Elements that define a model are:

➢ Architecture, which defines the structure of the network: the number of layers, the number of neurons in every layer and the way information flows on the network. There is no rule to determine the best architecture, so to get the right result we should test several architectures;
➢ Learning process, which specifies the process by which weights change in order to determine reports between inputs and outputs, or just between inputs. Learning process is divided into two main groups:

- Supervised way, which includes an external teacher, so that every output unit is told what should be the desired answer for input signals.
- Unobserved way, which does not include an external teacher and relies only on local information. We also refer to it as self-organization, in the sense that it organizes the data presented on the network by itself and reveals their common properties.

➢ Activation function, which produces neuron output based on inputs coming from other neurons.

Artificial neural network performance depends on the determining of the elements mention above.

Bayesian networks as well as other data mining are used to provide insight into an unknown area. They are structures for the representation of probability distributions concisely and clearly. Bayesian networks are used when the probability that an event occurs depends on the probability that a previous event has occurred. Their structure is a graphical structure, visually they are presented by means of a graph and in data mining this graph is directed and without cycles. Each node of the graph represents a feature of the data, while directed ribs represent probability dependence between relevant features. Bayesian networks make possible the calculation of the data feature that we want to predict the probabilities that this feature will take each of its values based on the probabilities that have the values of each of the other data features. One of the advantages of using Bayesian networks compared to other data mining techniques is the ease in handling missing data at the data for which we want to make predictions.

Decision trees show explicitly the reached decision (the forecast). The forecast is presented in visual form. These methods are structural descriptions of the decision taken presented in a very understandable format that shows the connections between different data features and the forecast. Visually they are diagrams in the form of a tree, they have a root, branches and leaves. Nodes (places where we have branching) represent tests about a data feature. A branch of the tree from root to leaf shows explicitly all their features and values that have influenced to the forecast (therefore the ones that have not influenced, too). The missing values create problems at the data for which we want to make predictions, because when the feature, whose value is missing, is tested, it is not clear on which branch it will continue.

Artificial neural networks are used when the data features, which are being processed are numerical in nature. Our data set contains numerical and nominal features. Therefore we cannot use artificial neural networks to work with the data collected. Bayesian networks provide the probabilities that the feature being predicting will take its possible values based on values that other features will take, whereas decision trees present in a clear way the decision that is taken (the forecast). Bayesian networks work better with missing data whereas decision trees do not.

Set of data that we have collected has no missing values because the questionnaire is formulated in such a way that all the fields be completed. We have also compared each methods performance by applying them practically on our data using weka environment. Weka is a free environment, that allows you to put into practice different methods of data mining on real data. The method of decision tree gives a better performance than the

method of Bayesian networks when applied on data that we have collected. Success rate that this method has in its predicting on the feature the level of knowledge, based on our data is 3% higher than the same prediction made by the method of Bayesian networks.

Based on the theoretical and practical comparison of forecasting methods we think that for the nature of our data, the method of decision tree is more appropriate to identify the problem in teaching and learning of Informatics in high schools. One of the best known algorithms, which generates a decision tree is algorithm ID3 created by Ross Quinlan. This algorithm uses the information gain criterion to determine which feature is going to be tested at every step and works with nominal data. Algorithm was improved by including procedures for working with numerical data, noise in the data, missing values and the generating rules by decision trees. The new algorithm called C4.5 was created by Quinlan, too. Precisely this algorithm we will use to get the necessary information from our data.

By responding to our survey questions, pupils and students gave us the information about their level of knowledge and interest for each of the curriculum areas of Informatics. According to the fields that have had the greatest impact on the level of knowledge and for which is displayed more interest from students it is designed the curriculum for the master of science in teacher of Informatics. The curriculum for the Master of Science in teacher of Informatics, in addition to basic training for teachers, it also includes all the subjects that provide or meet teacher's knowledge in different lines of Informatics.

The area that has the greatest effect on the level of the acquired knowledge in the high school is programming and algorithms. It is also the area that students consider as the most important in Informatics and where they have most of the difficulties. For this reason in the curriculum for master of sciences it is paid special attention to this area and has been decided to be 3 subjects (full course or module).

Figure 1 presents the knowledge received through the method of decision trees created by algorithm C4.5 that we used. Through this decision tree we get information about areas that affect or have affected more on the level of knowledge that students have acquired in high school in the subject of Informatics.
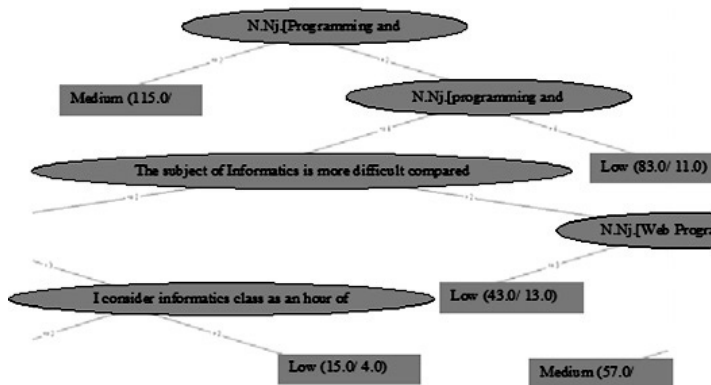

Figure 1. Features affecting more the level of knowledge.

Significant influence on the level of knowledge acquired in high school also has

the area Web Programming. (html, css, javascript) for which is decided to be a subject for the acquisition of basic knowledge and an optional subject for the implementation of this knowledge.

A subject is also devoted to architecture and computer organization, an area that also has a significant influence on the level of knowledge acquired in high school.
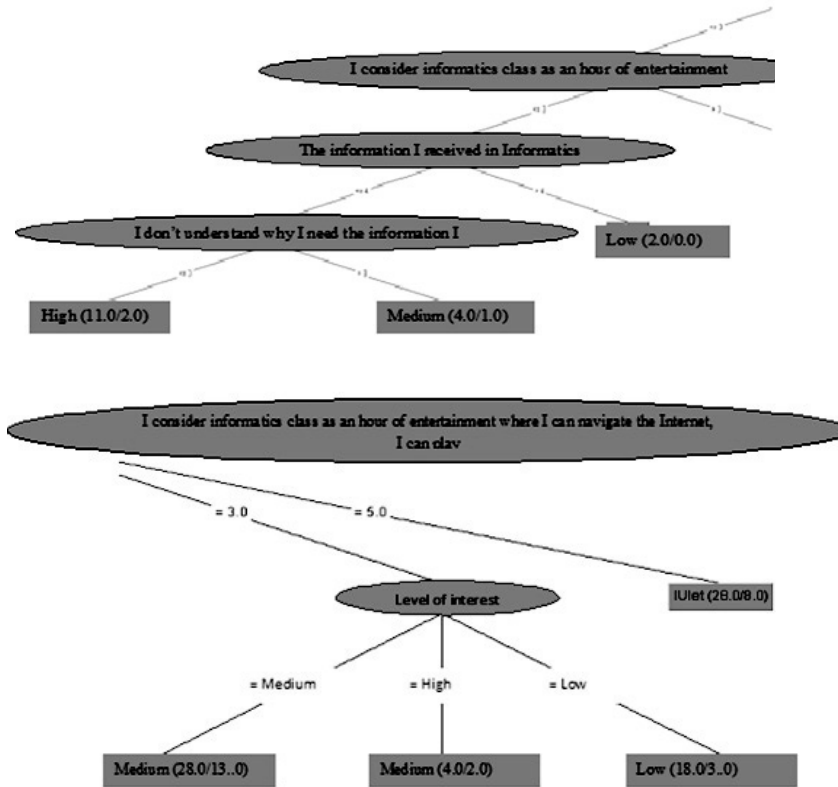

Figure 2. Impact of feature "Informatics as a leisure class" on the level of knowledge**.**

Pupils and students have also expressed the opinions they have about acquired knowledge in the subject of Informatics. Questions were as follows:

Rate from 1 to 5 how you adjust the following sentences regarding the assimilation of knowledge in the subject of Informatics.(1-certainly not,2-a bit,3-maybe,4-a lot, 5completely)

1. The subject of Informatics is more difficult compared with other scientific subjects such as mathematics, physics, etc.
2. I consider the class of Informatics as leisure activity during which I can navigate the internet, use the computer or play different games.
3. I don't get why do I need all the knowledge we receive in the subject of Informatics. Not everything is interesting to me.
4. I think that the knowledge that I am receiving in the subject of Informatics will be helpful to me in further studies or in work.
5. The information we receive during Informatics class, can be learned deductively.

Figure 2 presents the knowledge gained by the method of decision trees about the data features that affect more on the level of the knowledge gained by students during high school.

Those who consider Informatics class as leisure activity during which they can navigate the internet, use the computer or play different games have a lower level of knowledge than others, while those that do not consider it as such have an average or high level of knowledge.

To increase the interest of pupils and students in Informatics class it is intended that, in the curriculum for the master of science in teacher of Informatics, to have clause that make possible for the teachers of Informatics to recognize and to carry out independent research work in various matters to harmonize the knowledge acquired in each of the areas of Informatics with pedagogical experiences.

Pupils have been asked about the classes of Informatics developed in the laboratory. Pupils who have developed 50-100% of the classes of Informatics in the laboratory have mainly a medium level of knowledge and in a few cases when the level of their interest to learn the subject of Informatics is low they have a low level of knowledge. If they have developed 0% or 25% of the classes on the laboratory they have a low level of knowledge and in few cases medium level (in the case when 25% of the classes are developed in the laboratory and the interest level is low then they have a low level of knowledge, if the interest level is high than it is medium level and so on).
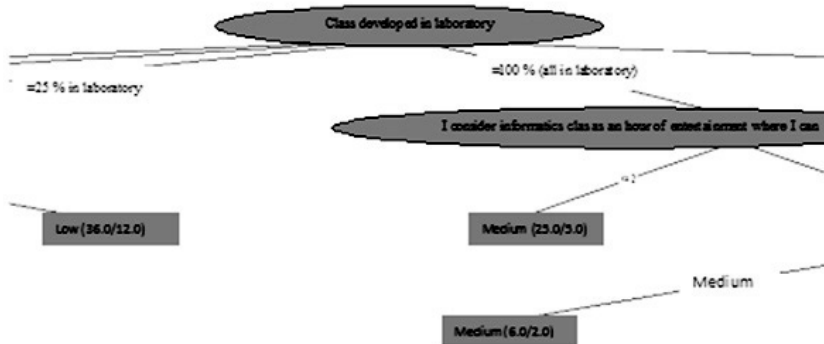

Figure 3. Impact of feature "class developed in the laboratory" on the level of knowledge.

Since the practice in each of the areas of Informatics has an impact on the level of students' knowledge, in the curriculum of each of the subjects is included the recognition and implementation of knowledge harmonization obtained in each subject by their use in practice.

Conclusions

Master of sciences that prepares the pedagogical staff for the Informatics subject in pre-university education is a necessity in Albania due to the incomplete knowledge that pedagogical staff of this field has in Informatics field. The Department of Informatics in Natural Science Faculty set up working groups to address the issues. The necessary data to identify the issues and address them were collected through interviews and questionnaires.

Adequate Data Mining technique for the nature of the collected data was determined to identify the problems faced during the course of Informatics. We analyzed the information obtained through data mining technique and as a result we created the master of science for informatics teachers in pre-university education.

Master of sciences besides basic education modules for teachers includes all modules that meet or provide teachers' knowledge in different areas of Informatics. The area that has more effect on the level of knowledge acquired in high school is programming and algorithms. Students consider it as the most important field in Informatics. They also consider it as the most difficult field of informatics. For these reasons we included 3 subjects (full course or module) in the curriculum of master of sciences in this area. Given that the field of Web Programming (html, css, javascript) has a considerable impact on the level of knowledge acquired by pupils, there will be two courses in the master of science for Informatics teachers related to this field. One of the courses will aim to provide basic knowledge in this field and a course will be optional for implementing this knowledge. A course is also devoted to the architecture field and computer organization, a field which also has considerable impact on the level of the knowledge acquired in high school.

To increase the interest of pupils and students in the informatics class it is intended for teachers that will attend this master to recognize and carry out independent research work in various matters that harmonize acquired knowledge in each of the areas of Informatics program with pedagogical experiences. Practice in each of the areas of Informatics has an impact on the level of knowledge of students in Informatics field. For this reason in the curriculum of each of the subjects is included the recognition and implementation of knowledge harmonization.

Master of Science "Teacher in Informatics for high schools" will enable that the teachers' knowledge in Informatics not to be limited in using ICT as a tool. They will take the basic concepts of Informatics and will transmit them to their pupils.

Future Works

The method used will serve as a starting point for studying the data that will continue to be collected in the coming years. The addressing of issues that will be identified in the future will lead to the improvement of curricula for Master of Science in teacher of Informatics.

The most appropriate collection methods will be studied also to find out possible groupings in the entirety of the collected data, which will be used to formulate personalized teaching methods.

Later on, report detection methods will be studied to identify connections and dependencies between the features of data such as: The literature used, geographic area, gender, laboratory base etc

Reference

- Institute of Curriculum and Training (2008) The program of information and communication technology for   grades 10,11,12,Tirana
- Institute of Curriculum and Training (2010) The program of information and communication technology for   grade 11, Tirana

- Institute of Curriculum and Training (2008) The program from module web design and dremweaver of the advanced ICT subject for grade 10, Tirana
- Institute of Education Development (2010) On the development of the new high school curriculum,Tirana
- Joseph P. Bigus (1996),Data mining with neural networks : Solving Business problems from application development to decision support. McGraw-Hill Companies
- Palit A.K., Popvic D.(2005) "Computation Intelligence in Time Series Forecasting" ISBN 1852339489
- Quinlan, J. R.(1986) Induction of Decision Trees, Machine Learning