IJCSC International Journal of Computer Science and Communication Vol. 3, No. 1, January-June 2012, pp. 159-164

COURSE RECOMMENDER SYSTEM IN E-LEARNING

Sunita B Aher[1], Lobo L.M.R.J.[2]

[1]M.E. (CSE)-II, Walchand Institute of Technology, Solapur, Solapur University India, sunita_aher@yahoo.com

[2]Department of IT, Walchand Institute of Technology, Solapur, Solapur University India

ABSTRACT

Now a day E-learning is becoming popular. E-learning is essentially the computer and network-enabled transfer of knowledge & skill. E-learning is also referred to as web-based education and e-teaching. Some examples of commercial systems are Blackboard, WebCT and Top- Class while some examples of free systems are Moodle, Ilias and Claroline.

In this paper, we propose the architecture for Course Recommender System & how the data flows through this system. This system predicts the best combination of subjects i.e. the subjects in which students are more interested. Here we use the learning management system such as Moodle to collect the data from student regarding their course choices.

Keywords: Moodle, Weka, classification, association rule, clustering algorithm.

1. INTRODUCTION

Educational Data Mining is an emerging discipline, concerned with developing methods for exploring the unique types of data that come from educational settings, and using those methods to better understand students, and the settings which they learn in [9].

Now a days, Moodle (modular object oriented developmental learning environment)is one of the most commonly used free learning management system which allows the creation of powerful, flexible and engaging online courses and experiences. These e-learning systems accumulate a vast amount of information which is very valuable for analyzing students' behavior. They can record any student activities involved, such as reading, writing, taking tests, performing various tasks, and even communicating with peers. They normally also provide a database that stores all the system's information: personal information about the users (profile), academic results and users' interaction data. However, due to the vast quantities of data these systems can generate daily, it is very difficult to manage manually [2].

Data mining or knowledge discovery in databases (KDD) is the automatic extraction of interesting patterns from large data collections. Some of the most useful data mining tasks and methods are statistics, visualization, clustering, classification and association rule mining. These methods are used to uncover new, interesting and useful knowledge based on students' usage data. These data mining techniques can be applied to some of the e-learning problems such as provide course adaptation and learning recommendations based on the students' learning behavior, and detection of atypical student's learning behavior.

The basic idea behind this project is to find the best combination of subjects which will lead to more effective learning in a particular stream. It will also help to understand the behavior of the student that is in which subjects, students are more interested.

The rest of the document is organized as follows. Section 1.1 provides a background of the related research fields covering a brief introduction about each. Section 2 describes the Literature Review. Section 3 discusses the architecture of proposed system. The Discussion & future work is presented in Section 4 followed by references.

1.1 Background

This research integrates issues from the research field of Moodle, WEKA, Information retrieval and Data mining (Classification ,clustering and Association Algorithm). The following subsections include a brief overview of these topics and their relation to the newly proposed methodology.

1.1.1 Moodle (Modular Object-Oriented Developmental Learning Environment)

Moodle is an open-source course management learning system to help educators create effective online learning communities. An organization has complete access to the source code and can make changes if need be. Its modular design makes it easy to create new courses, adding content that will engage learners and it is designed to support a style of learning called social constructionist pedagogy [2]. Moodle keeps detailed logs of all activities that students perform. It logs every click that students make for navigational purposes and has a modest log viewing system built into it. Log files can be filtered by course, participant, day and activity. The instructor can

use these logs to determine who has been active in the course, what they did, and when they did it. Moodle stores the logs in relational database MYSQL.

Moodle can be used in many types of environment such as in education, training and development, and business settings. Moodle is shown in figure 1.

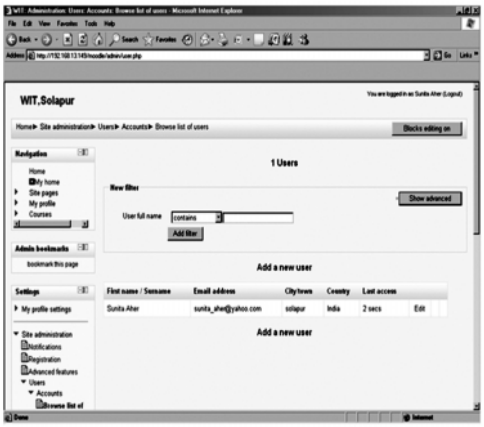

Figure 1: Moodle

1.1.2 Weka

The Weka workbench contains a collection of visualization tools and algorithms for data analysis and predictive modeling, together with graphical user interfaces for easy access to this functionality [7]. It is freely available software. It is portable and platform independent because it is fully implemented in the Java programming language and thus runs on almost any modern computing platform. Weka has several standard data mining tasks, data preprocessing, clustering, classification, association, visualization, and feature selection. The WEKA GUI chooser launches the WEKA's graphical environment which has six buttons: Simple CLI, Explorer, Experimenter, Knowledge Flow, ARFFViewer, and Log. Figure 2 shows Weka 3.5.3 with Explorer window.

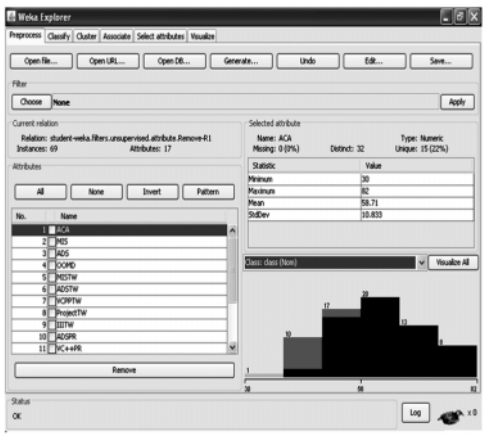

Figure 2: Weka 3.5.3 with Explorer Window

1.1.3 Information Retrieval and Extraction

Information retrieval (IR) is the science of searching for documents, for information within documents, and for metadata about documents, as well as that of searching relational databases from the Web. IR has the primary goal of indexing text and searching for useful documents in a collection. Research in IR includes modeling, document classification and categorization, user interfaces, data visualization, filtering etc. Information Extraction (IE) has the goal of transforming a collection of documents, usually with the help of an IR system, into information that is more readily digested and analyzed. IE aims to extract relevant facts from the documents while IR aims to select relevant documents. IE is interested in the structure or representation of a document.

1.1.4 Data Mining

Data mining is the process of discovering interesting knowledge from large amount of data stored in database, data warehouse or other information repositories. It includes various tasks such as classification, clustering, association rule etc.

1.1.4.1 Association Rule

Association rules are used to show the relationship between data items. Mining association rules allows finding rules of the form: *If antecedent then (likely) consequent* where *antecedent* and *consequent* are itemsets which are sets of one or more items. Association rule generation is usually split up into two separate steps: First, minimum support is applied to find all frequent itemsets in a database. Second, these frequent itemsets and the minimum confidence constraint are used to form rules. Support & confidence are the normal method used to measure the quality of association rule. Support for the association rule $X \Rightarrow Y$ is the percentage of transaction in the database that contains $X \cup Y$. Confidence for the association rule is $X \Rightarrow Y$ is the ratio of the number of transaction that contains $X \cup Y$ to the number of transaction that contain X. Association rule can be used in educational data mining for analyzing the learning data [7].

1.1.4.2 Classification

Classification is a data mining task that maps the data into predefined groups and classes. It is also called as supervised learning. It consists of two steps:

1. ***Model construction***: It consists of set of predetermined classes. Each tuple/sample is assumed to belong to a predefined class. The set of tuple used for model construction is training set. The model is represented as classification rules, decision trees, or mathematical formulae. This model is shown in figure 3.

2. ***Model Usage***: This model is used for classifying future or unknown objects. The known label of test sample

is compared with the classified result from the model. Accuracy rate is the percentage of test set samples that are correctly classified by the model. Test set is independent of training set, otherwise over-fitting will occur.

1.1.4.3 Clustering

Clustering is finding groups of objects such that the objects in one group will be similar to one another and different from the objects in another group. Clustering can be considered the most important unsupervised learning technique. In educational data mining, clustering has been used to group the students according to their behavior e.g. clustering can be used to distinguish active student from non-active student according to their performance in activities [7].

2. Literature Review

The paper [1] provides an up-to-date snapshot of the current state of research and applications of Data mining methods in e-learning. It provides the taxonomy of e-learning problems to which the Data Mining techniques have been applied including, for instance: Students' classification based on their learning performance; detection of irregular learning behaviours; e-learning system navigation and interaction optimization; clustering according to similar e-learning system usage; and systems' adaptability to students' requirements and capacities. In this paper they presents some tables which are very useful as e-learning material e.g. E-learning projects in which Data Mining techniques are used, e-Learning discussion forums, e-Learning organizations, Key e-learning research papers, Key e-learning books and books chapters, e-Learning information repositories, Learning objects repositories, e-Learning standards, Open source e-learning software etc.

This paper [2] describe the full process for mining e-learning data step by step as well as how to apply the main data mining techniques used, such as statistics, visualization, classification, clustering and association rule mining of Moodle data. Moodle does not provide visualization tools of student usage data; it only provides text information (log reports, items analysis, etc.).

In paper [4], Data Mining System (DMS) was designed and implemented to analyze the study records of two programming courses in a distance curriculum of Computer Science. Various data mining schemes, including the linear regression and probabilistic models, were applied to describe and predict student performance. The results mentioned in this paper indicate that a DMS can help a distance education teacher, even in courses with relatively few students, to intervene in a learning process at several levels: improving exercises, scheduling the course, and identifying potential dropouts at an early phase.

In paper [3], they propose proposed how to use adaptive machine learning algorithms to learn about the student's preferences over time. First they use all the background knowledge available about a particular student to build an initial decision model based on learning styles. This model can then be fine-tuned with the data generated by the student's interactions with the system in order to reflect more accurately his/her current preferences. The whole process is shown in Figure and is performed according to the following steps.

The paper [5] suggests the use of web mining techniques to build recommender system that could recommend on-line learning activities or shortcuts in a course web site based on learners' access history to improve course material navigation as well as assist the online learning process. These techniques are considered integrated web mining as opposed to off-line web mining used by expert users to discover on-line access patterns. They have proposed an approach to build a software agent that uses data mining techniques such as association rules mining in order to build a model that represents on-line user behaviors, and uses this model to suggest activities or shortcuts.

3. ARCHITECTURE OF PROPOSED SYSTEM

3.1 Component in Architecture

The architecture consists of following component as shown in figure 7:

3.1.1 Collect Data

The LMS (Learning Management System) can be used to collect the student's usage and interaction information. Here we use the Moodle system. Moodle is an open-source course management learning system where we add 13 categories and 82 courses. We create the login of each student on Moodle so that student can access the Moodle and give the choice regarding the subjects; he/she is interested in. The course categories are shown in figure 6. The activity chart for student is shown in figure 5.

3.1.2 Database

Moodle will store student's choices about the courses in Moodle database from where we can collect the data.

3.1.3 Data Preprocessing

The data is cleaned and transformed into an appropriate format to find the best combination of subject as the real world data is not suitable.

3.1.4 Result Using Data Mining Tool WEKA

The data mining algorithms are applied to discover and summarizes the knowledge of interest to the user i.e. to instructor, student and administrator. To do so, either a general or a specific data mining tool, or a commercial or free data mining tool can be used. Here we are using Weka, the open source data mining tool.

3.1.5 Build the Model

This is our module where we propose the algorithms which may be the combinations of various data mining tasks such as classification, clustering or association rule. We compare the result of these algorithms with existing available algorithm in Weka and find the best algorithm which will give you the best combination of subjects based on student's choice.

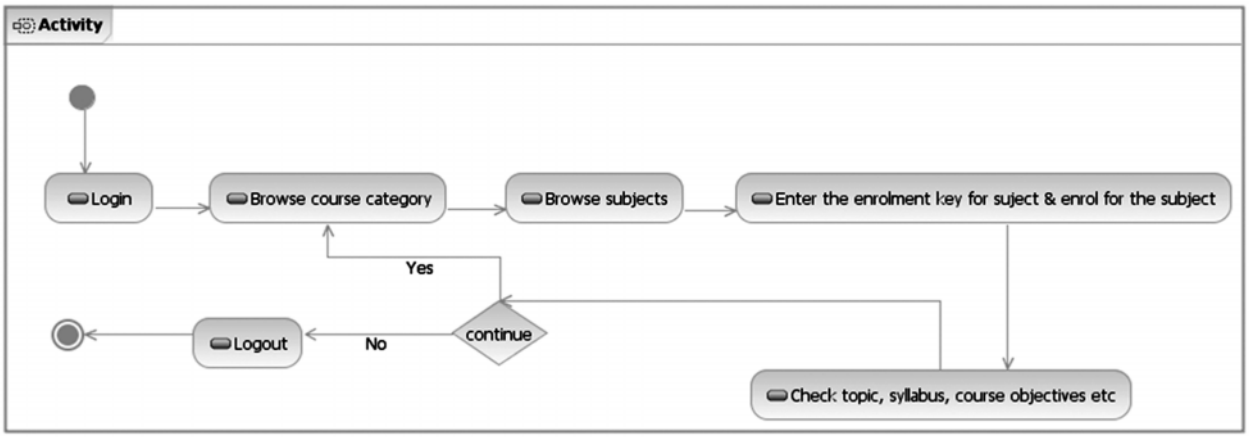

Figure 5: Activity Chart for Student

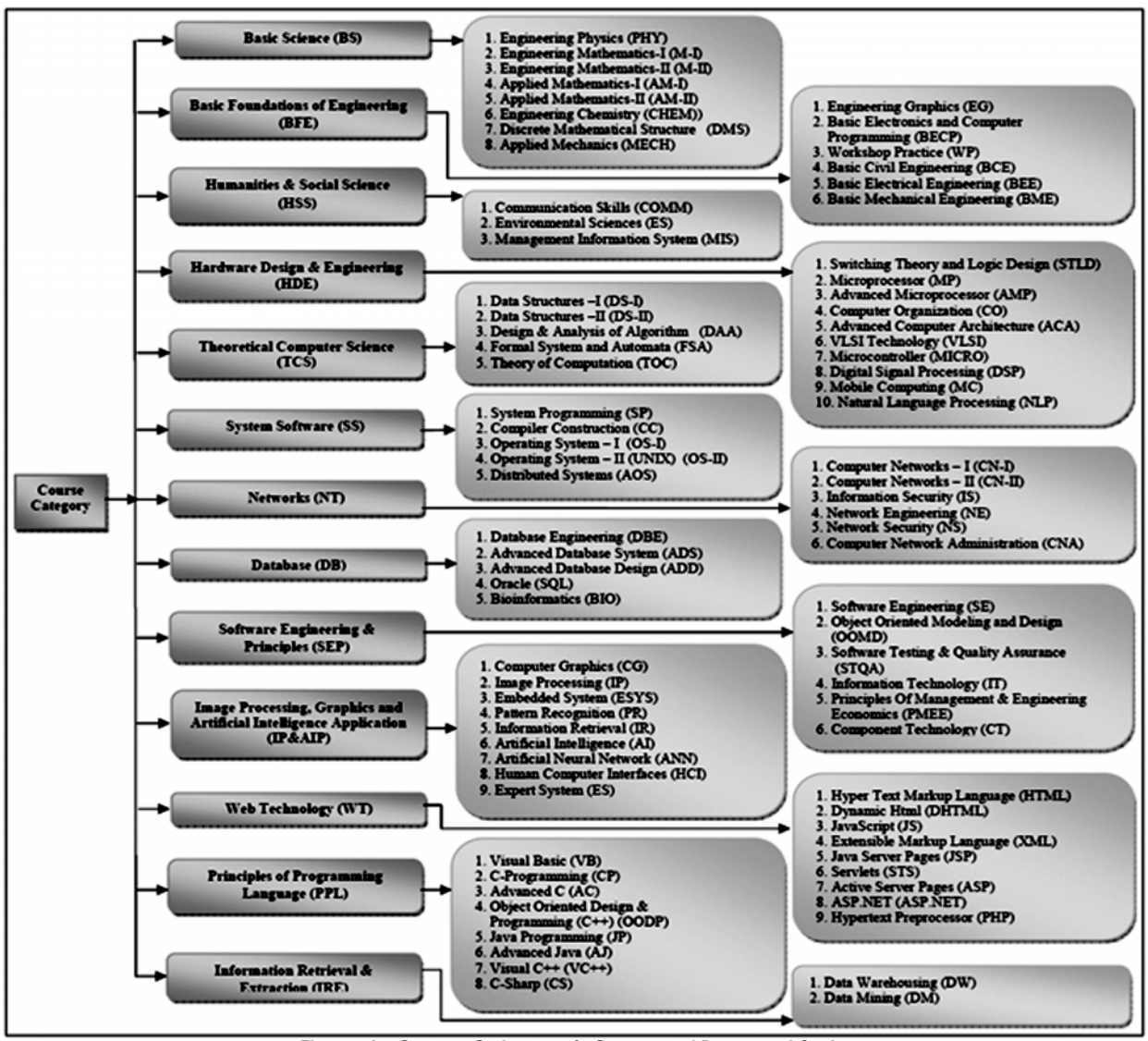

Figure 6: Course Category & Courses of Proposed System

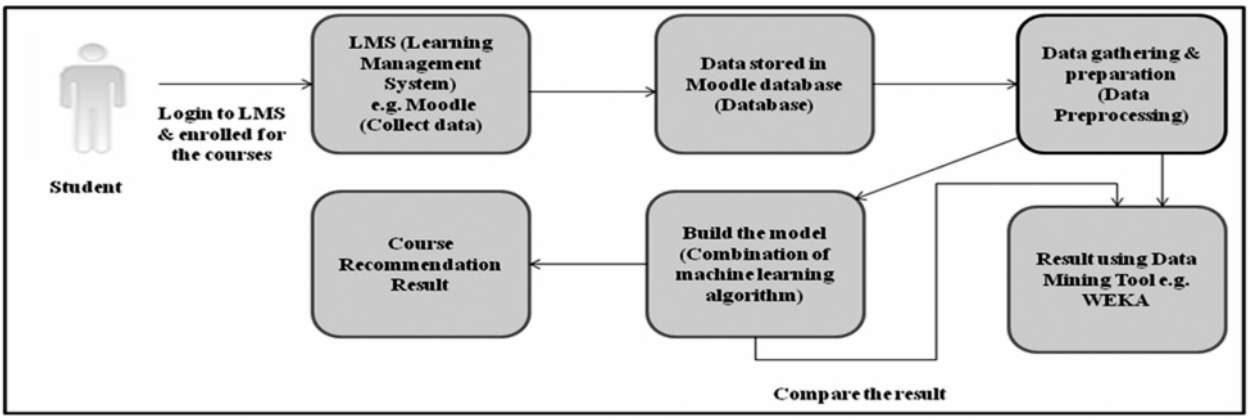

Figure 7: Architecture of Proposed System

3.1.5 Course Recommender Result

This module gives best combination of courses using our module which may be the combination of various data mining algorithm after comparing the result with WEKA.

Data flow diagram for this architecture is shown in figure 8.

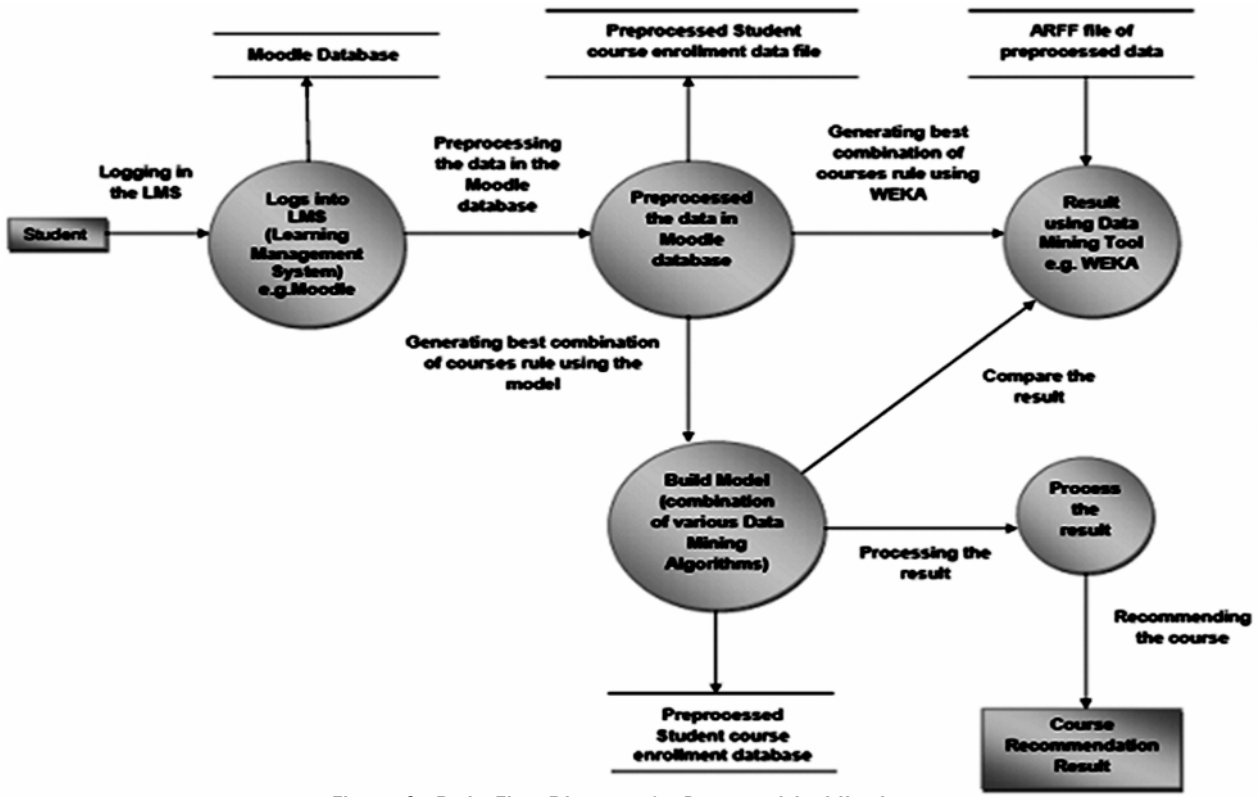

Figure 8: Data Flow Diagram for Proposed Architecture

3.2 Algorithm for Proposed Architecture

Algorithm for proposed architecture in pseudo code is given below

Step 1: Create Student Login in Moodle.

Step 2: Add the course category & courses under each category.

Step 3: Allow the student to logs in the system.

Step 4: While user logs in the system

- Allow user to view course category
- Allow user to view courses under each category
- Enroll for those subjects in which he/she is interested

Step 5: Select the data from Moodle database & analyze it.

Step 6: Preprocess the data obtained using Moodle Database

Step 6: Check the best combination of subjects result using the open source data mining tool WEKA

Step 7: Develop the algorithm which may be the combination of various data mining algorithm.

Step 8: Compare the result of this algorithm with those obtained using the WEKA

Step 9: Display the best combination of subject result

4. DISCUSSION AND FUTURE WORK

In this paper, we propose the architecture for Course Recommender System which predicts the course selection by the student. We collect the course choices of student using LMS (Learning Management system) such as Moodle. TO compare the result we propose to use open source data mining tool WEKA. This project would also be helpful to students in selecting the best combination of courses. This project would also be beneficial for selecting excellent course material for conducting staff development programs & workshops.

REFERENCES

[1] Castro F., Vellido A., Nebot A., and Mugica F. (in Press). "Applying Data Mining Techniques to E-learning Problems: A Survey and State of the Art. In L. C. Jain, R. Tedman, and D. Tedman (Eds.), Evolution of Teaching and Learning Paradigms in Intelligent Environment. *Studies in Computational Intelligence,* **62**, Springer-Verlag.

[2] C. Romero, S. Ventura and E. Garcia. "Data Mining in Course Management Systems: MOODLE Case Study and Tutorial". *Computers and Education*, 2007. Num. 51. pp. 368-384.

[3] C. Carmona, G. Castillo and E. Millán: "Discovering Student Preferences in E-learning", EC-TEL07, pp. 33-42, (2007).

[4] Hamalainen W., Suhonen J., Sutinen E., and Toivonen H. (2004). "Data Mining in Personalizing Distance Education Courses". *In World Conference on Open Learning and Distance Education*, Hong Kong (pp. 1-11).

[5] Za ï ane O. (2002). "Building a Recommender Agent for E-learning Systems". *In Proceedings of the International Conference in Education*, Auckland, New Zealand pp. 55-59.

[6] "Data Mining Introductory and Advanced Topics", by Margaret H. Dunham.

[7] Weka (2007). http://www.cs.waikato.ac.nz/ml/weka/

[8] http://docs.moodle.org/20/en/Main_Page

[9] http://www.educationaldatamining.org/

[10] Sunita B. Aher, Prof. L.M.R.J. Lobo: "Data Mining in Educational System Using WEKA".