A web-based intelligent report e-learning system using data mining techniques

Marija Blagojević *, Živadin Micić

Department of Information Technology, Technical Faculty Cacak, University of Kragujevac, Serbia

ABSTRACT

This paper presents a PDCA (Plan, Do, Check, Act) method of improving web-based intelligent reports of an e-learning system as intelligent system, which was created and implemented at the Technical Faculty in Cacak, University of Kragujevac. The focus is on improving LMSs (Learning Management Systems) or e-learning systems by predicting behavior patterns of students and adjusting the structure of these electronic courses. An existing learning management system is improved by using data mining techniques and increasing the efficiency of the courses using custom modules. This study presents the design, implementation, and evaluation of the system. Future work should relate to the continued improvement of the PDCA-created system, as well as the introduction of additional modules and a comparative analysis of the presented and future results.

1. Introduction

The expansion of e-learning has led to the increased use of systemic and continuous improvement of e-learning systems. This is also evident on the examples of Electrical Engineering teaching courses. The use of these systems has resulted in a need for monitoring and enhancing behavior patterns of all participants, with the aim of continuous improvement of the teaching process and ultimate results – education services. The paper presents the creation of a web-based intelligent report e-learning system using data mining techniques with PDCA (Plan, Do, Check, Act).

Learning Management Systems (LMSs), with numerous opportunities in the PDCA, have the ability to track and analyze user activity. Here, administrators can get reports on the activities of participants and statistical approaches at the level of each course (i.e. responsible teacher), as well as at the level of the entire system.

Each LMS has a database that contains records on the activities of each user. This characteristic of the system is very significant, with “plenty of information readily available, just a click away” [1, p. 2]. However, many of these records require a special tool for processing and extracting useful information. Such tools have limited capabilities and their use is mostly limited by the administrator’s choice, depending on the type of information. A universal solution for this problem lies in the use of data mining techniques, with the possibility of improving LMS [2,3]. Data mining or knowledge discovery in databases (KDDs) is the automatic extraction of implicit and interesting patterns from large data collections [4].

Part of the LMS report also requires improvement in terms of including web intelligence to detect significant patterns of behavior [5]. The need for such a solution includes intelligent and web-based aspects to meet the following requirements (in increments – the PDCA spiral):

* Reviews processed and approved for publication by Editor-in-Chief Dr. Manu Malek.
* Corresponding author. Tel.: +381 66005197; fax: +381 32342101.
E-mail address: marija_b@tfc.kg.ac.rs (M. Blagojević).

0045-7906/$ - see front matter © 2012 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.compeleceng.2012.09.011
• The prediction of behavior patterns.
• More interactivity.
• Visualization of the results obtained.
• More real time data analysis.

The remainder of this paper has the following structure. Related work is given in Section 2. Section 3 highlights the proposed framework and model, and Section 4 gives the purpose, tasks, and goals of the study. In Section 5, the methodology of creating the system, as well as the creation of dimensions, the OLAP (OnLine Analytical Processing) cube, data mining models, and the architecture of the system is presented. The implementation and evaluation of the system is given in Section 6. Finally, Section 7 highlights the improvement of the system characteristics, as well as its use on the examples of Electrical Engineering course. Concluding remarks are found in Section 8.

2. Related work

In recent years, many studies have used a standardized methodology for PDCA. In this paper, we aim to improve the LMS through the enhancement of courses by using data mining techniques. Data mining involves the automatic extraction of implicit and interesting patterns from large data collections [4]. Below, we present background on the PDCA spiral.

There are many papers that deal with the promotion of virtual learning environments, where the focus is on improving the PDCA spiral, and therefore the grouping of numerous works through PDCA. Frequently, innovations in the field of computer science are not readily available in the classroom [6]. In addition to introducing new systems into teaching, it is necessary to make the interaction with these systems secure and effective, but also fun [7].

Prakasam and Suresh [8] presented the architecture of an intelligent e-learning system that simplifies and automates the process of teaching in e-learning environments. One aspect that this system can offer is to monitor student progress. The main difference between the system in Prakasam and Suresh [8] and the system presented in this paper is that the data mining techniques applied in the former uses data on the web, while in our system, data mining techniques are applied only to data from the Moodle server. Gomah et al. [9] created an intelligent system to give students recommendations on the basis of the student model, whereas the system in this paper aims to make the recommendations regarding the use of the e-learning system available to the teacher. In addition to the aforementioned modules, the way students make recommendations in [9] describes the Tutor module, which determines what aspects of the domain knowledge should be presented to the learner.

Virtual learning environments in the form of LMS are becoming more prevalent in universities, providing various opportunities for the organization of blended learning, but also for the implementation of the entire online teaching process through the facilitation of lifelong learning [10]. These systems have been recognized as good pedagogical support for most activities, using less effort, time, and money, and not limiting the place from which the students can connect [8,11]. These systems provide numerous opportunities to work with the students, various activities with the possibility of collaboration, and knowledge gained through verification tests. Numerous commercial LMS have been developed, such as BlackBoard (BlackBoard, 2007) and WebCT (WebCT, 2007), as well as free systems such as Moodle (Moodle, 2007) and Claroline (Claroline, 2007). One of the most popular free systems is in fact Moodle [12–14].

The approach used in this study, which involves a combination of business intelligence and educational data mining, provides a successful solution in Aziz et al. [15]. In the study [15], an intelligent system identifies groups of students with greater and lesser capabilities based on behavioral pattern analysis. In addition, the system also optimizes the time it takes to perform current and historical data analyses. There are a number of works in which the PDCA approach assures quality in e-learning. In Walasek et al. [16], the PDCA approach was used to plan, describe, create, implement, and evaluate a number of online courses. Here, all documents were created in accordance with the Deming cycle, thereby providing high quality electronic courses [16]. The work of Santos et al. [17] shows the use of the PDCA process approach in the evaluation of electronic tests. Through the four Deming phases, the study [17] describes the procedure of evaluation, as follows:

- In phase P (Plan), four tasks are defined: establishment of participants, design of questionnaires, locating passive participants, and pilot tests.
- Phase D (Do) provides information on completing the questionnaire, forwarded schools, and specificity.
- In phase C (Check), analysis gives concrete results in regards to the percentage of completed questionnaires.
- Finally, phase A (Act) gives identified measures to improve the obtained results.

Similarly, Tanigawa et al. [18] used the PDCA approach for the improvement of teaching materials, which presented the design, implementation, and evaluation of electronic courses.

The PDCA model represents a foundation from which more complex models can be created, as suggested in Valkanos [19]. In Valkanos [19] a new model (ADDURI) is created based on the PDCA model. This model (ADDURI) consists of six phases, namely: Analyze, Design, Develop, Use, Improve, and Review.

The analysis of papers that deal with the application of the PDCA model to improve a component of or the entire LMS in relation to our study highlights some similarities and differences, as follows:

Similarities:
Our study achieves the improvement of existing systems through the phases of the PDCA model, as it is done in [16–18]. The PDCA model was used to improve the field of e-learning.

Differences:

- Our study focuses on a LMS in order to improve reporting, in contrast with the other studies [16–18].

3. Framework and proposed model

The Moodle LMS has the option of reporting, giving administrators and the creators of the course some insight into the functioning of the course and the whole system. The report that the teacher or the course administrator receives gives the possibility to adapt groups, participants, dates, and activities. A Moodle LMS administrator has more power using the field report. In this sense, the system administrator may receive a report for the complete system, except for certain courses. Based on a given SQL (Structured Query Language) query on the Moodle database, a report is created and can be downloaded in three different formats: text, ODS (Open Document Spreadsheet), and Excel.

The current system allows administrators of courses (i.e., teachers) to receive reports based on existing records in the database. However, there is a need for a reporting system that predicts user behavior based on existing patterns of behavior. This technique would require data mining. Fig. 1 presents the design of reporting systems and provides the possibility for prediction. Our study presents the Deming PDCA cycle [20], in creating web based intelligent report e-learning system.

Fig. 1 shows the reporting activity in a model system through our four activities (Plan, Do, Check, and Act).

4. Purpose, tasks, and goals (Step 1: P-Plan)

Details of the user’s activities cannot be predicted by the LMS. However, the LMS makes it possible to predict patterns of behavior. Therefore, there is a need for an intelligent, web-based system that can provide an interactive approach to predict user behavior patterns.

![Diagram of PDCA cycle](image-url)  
**Fig. 1.** Model design of the report system.
Our research goals in this study include the following:

- To identify features of data mining techniques to obtain predictions of user behavior patterns.
- To allow teachers to have access to future patterns of behavior for their students; based on this information, adjustments can be made in Moodle courses.
- To improve the existing system of reporting, such that the Moodle LMS will have web intelligence using data mining techniques.

Our research tasks include the following:

1. Defining the first formal structure of the environment and relationships within the e-learning system.
2. Creating appropriate data mining models.
3. Writing Data Mining Extensions (DMXs) to obtain predictive patterns.
4. Creating a web-based application that has the ability to interact with the user.
5. OLAP analysis: the designing of a multidimensional structure in which the main factors under analysis (i.e., year, month, day, time, minutes, course, activity, and module) will be taken as dimensions and will be used later to build an OLAP cube in order to analyze the recorded data.
6. Data pre-processing to clean and prepare the web server log file.
7. Describing the process and object models.
8. Selecting key resources for the Moodle LMS.
9. Creating links between applications and a DMX-over-Adomdnet.
10. Defining educational elements.
11. Visualizing the results.

5. Methodology of designing the system (Step 2: D-DO)

The system in this study was created based on logs from the Moodle database. The Moodle LMS has been used at the Technical Faculty in Cacak (www.tfc.kg.ac.rs) since 2007 in order to support blended learning. More than 100 courses have been developed within this system. Specifically, electronic courses have been created in the areas of Electrical Engineering and Information Technology. The number of active users, according to the system logs, amounted to 1935. The process of making an intelligent, web-based reporting system includes several stages, as outlined in the following subsections. The following subsections present chronologically ordered sub steps within D (Do) step.

5.1. Preprocessing

Preprocessing is a phase that involves removing entries that contain errors. Here, anything that is not relevant to this specific research is considered an error. In addition, time is presented in columns that include the year, month, day, hour, and minutes; this kind of presentation differs from the format originally obtained from the Moodle server log file. Columns that relate to additional information and to URLs are irrelevant in this research, and are therefore excluded from the log files.

5.2. Creating dimensions and the OLAP cube

After the pre-processing, the OLAP cube and dimensions (i.e., module, activity, course, year, month, day, hour, and minutes) are created.

Microsoft Visual Studio 2008 [21] is used for creating the OLAP cube. The data source and the data source view must be created before the cube itself is generated. The data here does not originate from an OLAP database; another database, a relational database, is the source. Fig. 2 shows the OLAP cube. After creating the initial data source, the data source view is generated. The data source view has two aims:

- It identifies the tables and views from the data source that will be used in the concrete Analysis Service Project, and
- It retrieves and stores the metadata about these objects, allowing for the building of cubes without having to maintain an active open connection to the data source.

After creating the data source and the data source view, the dimensions and cube are generated. The production of dimensions and cubes leads to the final phase of work in Microsoft Visual Studio 2008. This phase involves deployment solutions. Deployment solutions pass several phases, and the result is a successfully completed deployment in which the work is continued within the SQL Server Management Studio [23].

5.3. Creating data mining models

Data mining models were then created within the Visual Studio 2008 as follows.
5.3.1. Decision tree

Many studies from previous literature use the model decision tree to improve processes within the LMS e.g. [24,25]. In our study, a decision tree is used to predict the percentage of occurrence modules with selected input parameters. This is done through the DMX queries. The input parameters are hour, day, month, and activity. The DMX query is given in Section 5.4.1, and a query result is shown in Fig. 3.

5.3.2. Neural network

A number of studies deal with the use of artificial neural networks in the field of e-learning [26,27]. In our study, the information on the impact of all attribute values that relate to the predictable state is given. The Microsoft Neural Network algorithm uses a gradient method to optimize the parameters in multilayer networks in order to predict multiple attributes.

Input data are randomly split into two sets, a training set and a testing set, based on the percentage of data for testing and the maximum number of testing cases provided. The training set is used to create the mining model, and the testing set is used to check model accuracy. In this study, the percent of the testing cases is 30%.

The neural network model is illustrated with a visualization of prediction results in Fig. 4.

5.4. Writing the corresponding DMX query and test models

The SQL Server Management Studio accesses and checks the writing of DMX queries that are needed to obtain an adequate prediction of behavioral patterns. DMX query language is largely used for model creation, model training, prediction, and content access [28]. Fig. 4 shows the results of the executed queries. The request was written using the decision tree model, while the other two models were tested in Visual Studio, also presented in this chapter.

![Fig. 2. OLAP cube [22,30].](image1)

![Fig. 3. Results of the performed query.](image2)
5.4.1. Example of DMX query for decision tree model

\[
\text{select t\.}, \text{Predict ([Module]) as [Module],}
\]

\[
\text{vba!format (Predictprobability ([Module]), 'Percent') as [Probability]}
\]

\[
\text{from}
\]

\[
\{\text{dectree}\}
\]

\[
\text{prediction join}
\]

\[
\text{openquery}
\]

\[
\{\text{Moodle baza},
\]

\[
\text{select Hour, Day, Month, Activity from log'}\text{as t}
\]

\[
\text{on [dectree].Hour=t.Hour and}
\]

\[
[dectree].Day=t.Day and
\]

\[
[dectree].Month = t.Month and
\]

\[
[dectree].Activity = t.Activity
\]

This DMX-in predictable variable is a module whose probability is shown as a percentage. The Moodle database is a database in which records of the activities of the LMS are stored. There are approximately 1 million records. The input variables in this case are hour, day, month, and activity.

5.4.2. Neural network

A neural network is used to determine the values of attributes that favor and disfavor the predictable states. The viewer provides information on the impact of all attribute values that relate to the predictable state. The neural network model (Fig. 4) shows the most important attribute value that favors action, for instance, “forum” is an action with a value of “report”.

The results obtained from the neural network technique give a clear visualization of the predictable conditions, which may be important when teachers determine the concrete actions to be employed on an electronic exchange.

5.5. Architecture of the system

Intelligent reporting systems take records of user activities within the Moodle server. The data is processed and subject to execution in the corresponding DMX, depending on whether or not the user wants to see the predictive analysis. The results are displayed to the user within a web page and can then be executed during a new analysis. The system architecture is given in Fig. 5.

Generally, the architecture of this system can be viewed through its three main parts and the connections between them. The first part involves the SQL server and the relational database, which is preceded by preprocessing. The second part refers to the Analysis Server OLAP cube, including the processes that take place over it, and the third part deals with the web application. The link between the analysis and the web server applications is accomplished through ADOMD (ActiveX Data Objects Multi-dimensional) [29].
6. Implementation and evaluation (Step 3: C-CHECK)

For the evaluation system, two approaches are used: the implementation of a prototype system, and the implementation of the described cases.

6.1. Implementation of the system prototype

The system is implemented on a server for both the processing and preprocessing of the data, as well as the performing of data mining extensions. Given the number of records to be processed, which exceeds 1 million and is constantly increasing, the server should have above-average performance for faster response time. The implementation was successfully carried out, and then testing was performed using the described cases.

6.2. Implementation of the described cases

This section describes the case for a course in Electrical Engineering. Fig. 6 shows the selection of input data and the results obtained from the table. This case relates to a module system that predicts electronic modules within the course (module name and the associated probability).

Fig. 6 presents numerical parameters, i.e. the results pertaining both to the input data and the probability of the occurrence of modules within the e-course. The numerical result relating to the probability (ranging from 0 to 1) in this specific case amounts to 0.289. The system also provides the possibility of expressing the result in percentages (from 0 to 100).

The results given in Fig. 6 were obtained for a given time \( t \) (February 4th 2009, 00 h). By implementing the system through PDCA approach, an analysis was conducted for the same date and time in 3 more years (i.e. times \( t_1, t_2, \) and \( t_3 \)). These results are given in Table 1. For each of the aforementioned times, PDCA is implemented as follows: \( P_{t_1}D_{t_1}C_{t_1}A_{t_1} \), \( P_{t_2}D_{t_2}C_{t_2}A_{t_2} \), and \( P_{t_3}D_{t_3}C_{t_3}A_{t_3} \).

The results are analyzed in accordance with the iterations of PDCA model. The analyses are conducted at the beginning of the summer term (February 4th) by the “introduction” of courses in numerous fields, such as Electrical Engineering. The completion of the first cycle provides the results on the behavior patterns that enable prediction. In the next steps (by repeating the PDCA cycle in the ensuing time intervals) the obtained results extend the existing knowledge about user...
behavior patterns and are more precise than the previous ones because they include a greater number of records in the log file.

PDCA is implemented in spirals of increasing knowledge of the system that converge on the goal, each cycle closer than the previous.

The steps of PDCA cycle which recur in given time intervals are further explained in detail in the following sections (Step P (Plan) – Section 4, Step D (Do) – Section 5, Step C (Check) – Section 6 and Step A (Act) – Section 7).

7. Improving the characteristics of the created system (Step 4: A-ACT)

7.1. The main characteristics of the system created

The main characteristics of the created system are as follows:

- **Simplicity**: The architecture and use of the system are in general simple. The interface is user-friendly and requires no special preparation to use it.
- **Enabling more real time data analysis**: Given the fact that the database is constantly updated with new records of user activity, the application automatically performs the same process, based on real time analysis.
- **Cost effective**: This system is cost effective. The only required costs are related to the server for data storage and processing. Considering the number of users on the Moodle system, as well as teachers who use the intelligent system, this cost is manageable.
- **Web intelligence support feature**: The system includes a web intelligence related to prediction, which is the main purpose of the system (i.e., obtaining predictive behavior patterns).

The proposed system, unlike the existing one, has additional possibilities for improving e-courses, as well as the teaching process, through the prediction of user behavior patterns. Table 2 gives a comparative review of the properties of the current – “existing system” and the newly created “proposed system” intelligent web-based report e-learning system.

7.2. A possible scenario of use (courses in Electrical Engineering)

Moodle electronic courses give teachers insight into the basic activities of its users and students. However, material innovation and new activities are typically required each school year. Given the present and predicted patterns of student
behavior (in terms of time and amount of access), the teacher can create a course accordingly and increase the efficiency of the course.

For example, for courses in Electrical Engineering, the teacher can choose his/her own courses and an analysis technique according to the teacher’s interest. If a teacher creates a new course or updates an existing one, he/she should consider the form of teaching materials and the activities within the course.

The teacher enters the input parameters that are relevant to the analysis (e.g. ID of the course, module name, name of activity, and time). Based on the input data, the intelligent reporting system provides a table and a graphical representation of the results. The teacher receives display modules and activities associated with the percentage probabilities. The teacher can decide on the selection of modules and the organization of activities for students in the course depending on the percentage results. If the highest percentage is allocated to collaborative modules (i.e., forum, chat, wiki) this can then be arranged through the activity.

Using intelligent web-based reporting systems facilitates the process of policy-making and increases the opportunity for more efficient use. After the course is adjusted, a new PDCA spiral of quality begins.

8. Conclusion

A web-based intelligent report e-learning system created with data mining techniques leads to a modern and intelligent way of reporting user activity. Compared with the current reporting system in Moodle LMS, the proposed system presents an improvement since it predicts behavior patterns thus leading to the increased efficiency of the participants.

Advantages of the proposed system relate primarily to the opportunities provided by the intelligent features, while a disadvantage is the need for above-average performance of the server. The advantages are reflected in the existing web intelligence and the possibility of predicting user behavior patterns. Furthermore, these advantages have a direct effect on quality assurance in e-learning and on the improvement of the teaching process through the adaptation of content by predicting behavior patterns.

Through the Deming PDCA cycle of activities, the following can be concluded (in time $t$, $P_tD_tC_tA_t$):

- $P_t$: This includes the review of existing reports, and the need for identified reports that provide predictions of user behavior in an LMS. The conclusion at this stage involves moving towards recommendations for planning in terms of the need for inclusion of all relevant factors.
- $D_t$: This includes the design and implementation of the new system and it is consistent with the planned outcomes, adjustment for all activities, and successful functioning of an intelligent, web-based reporting system.
- $C_t$: This is the compulsory phase control, carried out through the testing activities, installation and evaluation systems, the execution of change, and the improvement system. This stage also fulfilled its purpose, and the changes are related to the visualization of results in order to improve them.
- $A_t$: This includes the possible improvement through the re-use of system analysis tasks. Here, future work should be related to the creation and implementation of new modules in the given time $t_n$, $P_nD_nC_nA_n$.

The approach presented is similar to some previous. However, this study gives improvement in terms of the report system in the field of e-learning. Moreover, development and implementation of new modules, as well as user authentication, will be the subject of future research.

Acknowledgment

This study was supported by the Serbian Ministry of Education and Science (Project III 44006, http://www.mi.sanu.ac.rs/projects/projects.htm#interdisciplinary).
References


Marija Blagojević was born in 1984 in Cacak, Serbia. She graduated from Technical Faculty Cacak, University of Kragujevac (B.Sc. 2007) where she is currently a Teaching Assistant (Department of Information Technologies). At the moment she is a Ph.D. student and her areas of interest are e-learning and data mining. She wrote several articles about web mining application in e-learning and detecting behavior patterns in online environment. Recently, she has studied on improving learning management system with addition of web intelligence.

Živadin Mićić was born in 1955. He received B.Sc., M.Sc., and Ph.D. degrees in Mechanical Engineering in 1979, 1982, and 1997, respectively. His field of specialization is Information Technology and Systems. Since 1999 he has been a Full Professor at Technical Faculty Cacak, University of Kragujevac, acting as Head of the Department of Information Technology as of 2001. He is the author of over 220 papers, “Information Technology” monograph, “IT in integrated systems”, etc., especially teaching publications for online usage.