Analysis of Students’ Behaviour in the Web-based Distance Learning Environment

ZOLTÁN BAŁOĞH, MICHAL MUNK, MILAN TURČÁNI
Department of Informatics, Faculty of Natural Sciences
Constantine the Philosopher University in Nitra
Tr. A. Hlinku 1, 949 74 Nitra
SLOVAKIA
zbalogh@ukf.sk, mmunk@ukf.sk, mturcani@ukf.sk

Abstract: - One of the useful methods of increasing the effectiveness of the learning process and quality of students’ results is integrating the on-line content and learning management system. In the paper we describe the applicability of different types of resources and activity modules in the e-learning courses and the worthiness of their usage. The contribution is aimed at analyzing students’ behavior in the virtual learning environment, while the used source of data combines the ones on the system utilization obtained from the log file and the ones obtained using statistic finding.

Key-Words: - Web-based Distance Learning, e-course, usage analysis, e-learning, LMS, virtual learning environment

1 Introduction

During recent years designing and implementing web-based education (E-Learning) systems have grown dramatically [1] and this type of education is playing an important role in teaching and learning [2]. It is implementing as a new method of training which complements traditional methods [3] and its final ambition is to build an advanced society for citizens and support creativity and innovation [4].

In today’s active era of ICT usage not only in education but also in the field of industrial and automation equipment there is a question coming into the foreground: how should some processes be used and controlled both to have their course reliable and to achieve the demanded result [5].

With the rapid advance of the Internet, e-learning systems have become more and more popular [6],[7]. An e-learning system provides the following functions: (1) delivery of learning content for students via the Internet; (2) record of learning progress and portfolio; (3) management of learning content, assessment and course; and so on.

With help of a virtual learning environment teachers can make students’ study more effective and time-efficient. Carefully prepared multimedia study material placed into a virtual learning environment enables to demonstrate and visualize the subject matter more clearly and comprehensible, to develop students’ logical thinking, increase their imagination and help them to solve various problems [8].

The Internet and related web technologies do offer great solutions for presenting, publishing and sharing learning content and information, as is the case in many other areas. Special software called Learning Management System (LMS) is generally used in most institutions providing web-based learning [9], [10]. Nowadays, various LMS are used as a supporting tool in electronic education [11], [12]. The most of universities combine form of learning using one of a number of commercial or free LMS. They decided to use products such as Claroline, Fle3, ILIAS, MS Class Server, WebCT, Eden, Enterprise Knowledge Platform, LearningSpace, eAmos, eDceo, Uniforms, uLern, Aspen, Oracle iLearnin, NETOL School and Moodle [13].

1.2 Methodology of research

Our aim was to evaluate the course quality using various approaches. We used the following research methods.

Entrance questionnaire
It served to get the non-anonymous information about the statistical sample from the point of view of age, employment and practice.

Final (output) questionnaire
It provided the feedback about the quality of content, tests, used means, process of study, method
and effectiveness of study as well as the attitude of the participants toward e-learning.

The outputs of non-anonymous questionnaire investigation usually tend to be very subjective, which was the reason why we decided to rely not only on the participants’ responses but to find also some other, more objective point of view at the course modules. We were interested in how the users studied in the course, what was their navigation (the transition between the modules), which materials they accessed (eventually the number of accesses) etc. We needed to create the model of user’s behaviour in the particular course.

Usage analysis

Usage analysis (focusing on end-user behaviour) provides an influential second source of information. Very interesting and useful course usage information were gained also from the log-on file analysis.

A log file is an electronic file generated by a software package. A log file consists of the registered actions of end users in a predefined format. The minimum requirements for log file analysis consist of who, when, what and how. [14]. So it is a sequence of behavioral data (in our case recorded during the participants’ study), stored on a permanent medium [15].

Log file analysis is the systematic approach to examining and interpreting the content of behavioral data. Its goal is to assist in finding patterns in the behavior of people as they interact with a computer. [15] Analysis of log file served to formulate the association rules of participants’ behaviour in the e-course as well as the sequence and frequency of electronic sources accesses. Log file analysis is an important instrument to make the behaviour of these end users transparent [14].

That can help us better understand the behaviour of the student in the e-learning environment. During the data preparation we took into account recommendations resulting from series of experiments examining the impact of individual steps of data preprocessing on quantity and quality of extracted rules [16][17][18]. Analysis of log-on files is method of data mining. Data mining is a process that is used to identify hidden, unexpected pattern or relationships in large quantities of data. Data mining predicts future trends and behaviors [19]. Data mining scours databases for hidden patterns, finding predictive information that experts might overlook because it falls outside their expectations. In our case, the log-on file is created by LMS Moodle automatically and contains information of everything that happens in the e-course. In order to analyze the data we have to prepare them in advance – e.g. delete invalid data, convert system time, create categories of actions, etc. After the necessary data preparation we are able to find the model of user’s behavior in particular e-course - we can follow his navigation through the course, see what materials were visited and/or the time spent in the course. However, we cannot know if the opened material was really read and understood [20], [21].

2 Usage analysis

2.1 Course parts categories visit rate analysis

In the following part we are going to describe the results of association rules analysis, which represents a non-sequential attitude to the data being analysed. We shall not analyse sequences, but transactions, i.e. we shall not include the time variable into the analysis. In our case transaction represents a variety of visited categories of the course by one user. Regarding our data, we shall consider one transaction to be the categories of parts of the course visited by one user for an observed period of time.

![Web graph – visualization of the found rules](image)

Fig. 1: Web graph – visualization of the found rules

The web graph (Fig. 1) [22] visualizes the found association rules, particularly the size of the node represents the support of an element, the line-width - the support of the rule and the brightness of the line - the lift of the rule. We can see from the previous graph, which clearly describes the chosen associations, that among the most frequently visited categories of the parts of the course belong: main page, quiz selftest, forum, practice assignment,
report a feedback entrance output (support > 80%), similarly as combinations of pairs of these categories (support > 70%) or, for example, that the categories of parts of the course - study material and help – occur more frequently jointly in the sets of visited categories of parts of the course by individual users than separately (lift = 1.9). The same applies to the categories - upload and help (lift = 1.2). In these cases the highest rate of interestingness was found (lift), which specifies how many times more frequently the visited categories occurred jointly than in case if they were statistically independent. In case that the lift is higher than one, the selected couples occur more frequently jointly than separately in the set of visited categories of web sections by individual users. However, it is necessary to become aware of the fact that upon characterizing the rate of interestingness – (lift), the orientation of the rule makes no odds. In case of the remaining found rules the value of the lift was approximately one.

The only requirement (validity assumption) of the use of chi-square test is high enough expected frequencies. The condition is violated, if the expected frequencies are lower than 5. The validity assumption of chi-square test in our test is violated. This is the reason why we shall not prop ourselves only upon the results of Pearson chi-square test of independence, but also upon the value of calculated contingency coefficient and graphic visualization of dependency.

Contingency coefficient represents the degree of dependency between two nominal variables. The coefficient value (Table 1) is approximately 0.3, while 1 represents perfect dependency and 0 means independency. There is a medium dependency between the number of accesses into individual categories of parts of the course and time periods of study, and the contingency coefficient is statistically significant. The zero hypotheses (Table 1) are rejected at the 1 % significance level, i.e. the number of accesses to individual parts of the course (Category) depends on the period of study (Term).

Table 1
Analysis of crosstabulation - Category x Term

<table>
<thead>
<tr>
<th></th>
<th>Pearson Chi-square</th>
<th>df</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Category x Term</td>
<td>562.8561</td>
<td>18</td>
<td>0.0000</td>
</tr>
<tr>
<td>Contingency coefficient</td>
<td>0.308845</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Source: own research

The graph (Fig. 2) visualizes interaction frequencies - Category x Term. The graph represents categorized polygon, where on the x axis are periods of study (Term) and on the y axis are observed frequencies, while one polygon is drawn for each level of the variable Category. If the curves copy themselves, they show equal course, and the answers are independent. On the other hand, if there is any rate of dependence, the curves would not copy themselves – they would have different courses. In our case the curves do not copy themselves, they have different courses – which only proves the results of the analysis. Other course is observed in categories quiz selftest and forum.

Fig. 2: Interaction Plot - Category x Term

2.2 Extension of the data source with the data obtained by one’s own survey

Extension of the used data source containing data on the system usage obtained from the log file of the virtual learning environment with the data obtained by one’s own survey allows us to investigate utilization of individual activities of the course depending on the found characteristics of participants.

Table 2
Analysis of crosstabulation - Category x Age

<table>
<thead>
<tr>
<th></th>
<th>Pearson Chi-square</th>
<th>df</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Category x Age</td>
<td>319.1152</td>
<td>18</td>
<td>0.0000</td>
</tr>
<tr>
<td>Contingency coefficient</td>
<td>0.237507</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Source: own research

Using the one’s own survey, the age of participants – course users, the length of practical experience in the field and the way of their education, where we differentiate self-learning – self-studying and further learning by means of courses offered by the university. We suppose that utilization of individual
activities in the course depends on the following factors, such as age, practical experience and way of their further education. There is a minor dependency \( C = 0.24 \) between the number of accesses into individual categories of course parts and the age of participant, while the contingency coefficient is statistically significant. We refuse the zero hypothesis (Table 2) with a 99% reliability (it is rejected at the 1% significance level), i.e. utilization of individual course activities (Category) depends on the age of user (Age).

Interaction frequencies Category x Age are visualized in fig. 3, where age categories of participants – course users (Age) are presented on axis x and frequencies observed are presented on axis y, while for each level of the variable Category is depicted one polygon. Curves are not identical and have different courses – which only proves the results of analysis. Unlike course can be observed for categories quiz selftest, forum and study material. Older participants use activities quiz selftest and forum more frequently, while on the contrary, younger ones use practice assignment and study material above all.

Table 3
Analysis of crosstabulation - Category x Practice

<table>
<thead>
<tr>
<th>Category x Practice</th>
<th>Pearson Chi-square</th>
<th>df</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>659.0331</td>
<td>27</td>
<td>0.0000</td>
</tr>
<tr>
<td>Contingency coefficient</td>
<td>0.331501</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Between the number of accesses into individual categories of the course parts and the length of participants practical experience is mean dependency \( C = 0.33 \), contingency coefficient is statistically significant. Zero hypothesis (Table 3) is rejected at the 1% significance level, i.e. utilization of individual activities of the course (Category) depends on the length of practical experience of the participant (Practice).

Interaction frequencies Category x Practice are visualized in fig. 4, where categories of the length of practical experience of course users (Practice) are presented on axis x and frequencies observed are shown on axis y, while for each level of the variable Category is depicted one polygon. Curves are not identical and have different courses – which only proves the results of analysis. Unlike course can be observed for categories practice assignment, quiz selftest and study material. Participants with longer practical experience use the activity quiz selftest more frequently, while on the contrary, younger ones use practice assignment above all.

Table 4
Analysis of crosstabulation - Category x Study

<table>
<thead>
<tr>
<th>Category x Study</th>
<th>Pearson Chi-square</th>
<th>df</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>295.7249</td>
<td>9</td>
<td>0.0000</td>
</tr>
<tr>
<td>Contingency coefficient</td>
<td>0.229111</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Between the number of accesses into individual categories and the method of further education is slight dependency \( C = 0.23 \), contingency coefficient is statistically significant. Zero hypothesis (Table 4) is rejected at the 1% significance level, i.e. utilization of individual
activities of the course (Category) depends on the method of their further education (Study).

Interaction frequencies Category x Study are visualized in fig. 5, where values of the variable Study are presented on axis x, where we differentiate self-education – selfstudy and further education using the course offered by university, and the frequencies observed are shown on axis y, while for each level of the variable Category is depicted one polygon. Curves are not identical and have different courses – which only proves the results of analysis. Unlike course can be observed for categories practice assignment and study material. We assume that participants, who use courses offered by university for their further education, have experience with virtual learning environment and use individual course activities more frequently. On the contrary, participants, who use self study as further education, use mainly basic course activities such as practice assignment and study material.

3 Discussion

In most cases the usage analysis is applied to extract patterns / behavior patterns of web users [22]. In this article, we have focused on finding associations between the uses of the course activities.

The course was mainly used for the communication among the participants as well as the tool for the assignment distribution and collecting. Students were motivated for active studying also by the allotment of 20 credits for successful passing of this course necessary for their career advancement (the assignments were compulsory for the students to obtain this credits necessary for their carrier advancement). Among the most frequent moves from the main page we can name displaying the list of assignments (Assignment view all), displaying the list of users (User view all) and displaying of the list of test (Quiz view all).

Study materials were used much less than we expected. However, that does not mean that the students would not read them. It is possible that they printed them at their first display of them. On the other hand, students used tests quite a lot as they seem to be a good preparation for the state exams. The quiz reports show that a lot of students repeated the quiz attempts several times till they gained 100%. The outcomes of self-tests were displayed for some time for all the students so they could compare the results which increased the competition among them.

We have used data obtained through its own research to extend the logfile with a set of other variables such as age attendant, length of practice, the method further education and so on. It allowed us to monitor the use of activity rates depending on the variables selected by the questionnaire. Between the use of assets and monitored exchange rate variables were found only low to moderate interdependence. But this relationship was statistically significant in all cases.

4 Conclusion

The quality and success of the e-learning support of the study does not consist in the quantity of created courses (usually filled into a specified matrix on a mass scale). Success of the e-learning support of the study requires a systemic planning, creation of a draft, assessment and putting the system into practice, in which education is actively supported and stimulated. In order that the e-learning system was successful, it has to have a taste for all participating components, including learners, teachers, supporting personnel and the institution as well. So if we want to evaluate the quality of course, we have to take a look also on activities of tutors during study, not only on content (number of teaching aids, tests etc.)[22].

Via analysis like the one described above we will be able to make the e-courses more effective and attractive for the students to perform better effectiveness of e-learning study. It would be interesting to compare the questionnaire responses and the usage analysis separately for particular participants – that may be the objective of another research.
Acknowledgments. This publication is published thanks to the financial support of the project ESF 26110230026: A-CENTRE of the FNS, CPU in Nitra, Centre of Innovative Education.

References: