

Abstract

At the Budapest University of Technology and Economics (BUTE) in the Institute of Applied Pedagogy and Psychology (APPI) we started to use Moodle open source course management system in 2006. Formerly we served the curricula by using workbooks, multimedia CD-s, course-books, but due to the fast change of curricula it was a hard task to keep the hard copy of learning materials up-to-date. Later students could download learning materials with the help of using Internet, they could resort to the help of tutors instead of visiting the lessons. Blended-learning method is one of the most effective form from a number of our full-time gradual courses. Today we are using Moodle 1.9 version to support learning and training activities. Nowadays, we use questionnaires to evaluate users' satisfaction. These deliverances take an important place in the development of the content and appearance of our e-courses and curricula. We implement web usage mining as an appropriate tool for uncovering those parts of the user activities that cannot be told by the user personally. With the analysis of the collected web server log files we can track users' actions and by identifying typical student behaviour and matching these data with the given grades, successful learning strategies can be identified. Our first results show that qualitative and quantitative approach should be used parallel and with the help of data mining methods we can reveal more insights about the quality and usability of e-learning systems as we could before.

Keywords

e-learning · usability · data mining

1 Courses in an electronic environment

In Hungarian higher education the use of CMSs has become more and more popular even in full-time graduate courses. In our country, open source LMS (learning management system) and CMS (course management system)¹ systems are especially prevailing. Because it is a cheap solution, it offers practical keys for teachers, and it is easy to learn, use and operate the systems.

However, this does not mean that the environment and the contents of a system supporting learning processes can be used in an adequate manner. In spite of the efforts, we often created cluttered interfaces that are uneasy to use, and to oversee. A greater problem is that these malformations obstruct our main objective of the transmitted information and cognition to reach its aim.

2 Criteria of usability in an electronic learning environment

The qualitative aspects in connection with educational software products in accordance with Barker's and King's (1993) terminology are the following [1]:

- engaging interest
- interactivity
- tailorability
- appropriate proportion of the multimedia mix
- means and style of interaction
- quality of interaction
- quality of end-user interfaces
- suiting various learning styles
- monitoring and assessment techniques
- built in/integrated intelligence
- adequacy of ancillary learning support tools
- suitability for single user / group distributed use

¹ <http://moodle.org/>

Erika Jókai

Department of Ergonomics and Psychology, BME, 1111 Budapest, Egry J. u. 1., Hungary

e-mail: jokaie@erg.bme.hu

Electronic learning and teaching environments can be subjected to the same criteria. From the aforementioned criteria, engaging interest is an educational-didactical aspect, while the quality of the user interface and interaction is mainly a question of software ergonomics. The focus of our research is on interactivity, means and quality of interaction, and the identification of learning habits.

E-courses are often criticized emphasizing the high rate of dropouts, and the impersonal nature. During our analysis of e-courses used in higher education a high rate of dropouts has not been experienced, since gaining the scores for the course is motivated by acquiring the credit points as well. During the independent learning period there is no actual interaction between the student and the teacher, and between the students. Because of this, it is expedient to make the electronic environment more personal for example by creating a social network – with the use of forums, groups, chat systems, wikis, and with the integration of face-to-face sessions (practical activity, personal consultation).

3 Usage and usability

There can be many questions raised concerning the usability of the electronic learning environment: Who uses it? What for? When? How? With what results? Answer these questions somehow implies examining usability.

Usability testing is the process by which the human-computer interaction characteristics of a system are measured, and weaknesses are identified for correction. Such testing can range from rigorously structured to highly informal, from quite expensive to virtually free, and from time-consuming to quick. There are three main styles of testing [2]:

Exploratory testing examines a system and looks for areas of user confusion, slow-down, or mistakes. Such testing is performed with no particular preconceived notions about where the problems lie or what form they may take. The deliverable for an exploratory test is a list of problem areas for further examination. Exploratory testing can be used at any point in the development life cycle, but it is most effective when implemented at an early stage and several times.

Threshold testing measures the performance characteristics of a system against predetermined goals. This is a pass/fail effort. Threshold testing typically accompanies a beta release.

Comparison testing measures the usability characteristics of two approaches or designs to determine which suits better users' needs. This is usually done at an early prototyping stage.

Measuring usability is rather time consuming, the quality of the user interface, and the contentment of the users can be measured by applying methods of software ergonomics (e.g.: professional methods – guideline reviews, GOMS, cognitive admittance, heuristic analysis, empirical methods – enquiring the

users by for example questionnaires, personal enquiries, monitoring user activity, card sorting, focus group method, INTERFACE² analysis). However, monitoring system data concerning the use of the environment could also be adequate for this matter. These data can be used as a secondary product of e-courses for evaluation purposes. At an early stage databases and log files stored on the server were used to eliminate difficulties of access for the students – so we were made to get to know our courses better from an IT point of view as well. The identification of individual inconveniences has presented the opportunity of disclosing the general problems as well. We are trying to exploit this opportunity by data mining.

4 Data mining

The use of questionnaires has the potential limitation of the users only telling us what they are asked and thereof what they can remember. The log files of the web servers contain among other things the users' interactions with the system and with the elements of the curricula and by analysing them albeit is possible to identify user habits.

There are existing data mining software like for example the SPSS Clementine 12.0, which is used at the Department of Ergonomics and Psychology, BME in parallel with SPSS Web Mining for Clementine data ingesting node, that are used for analysing log files. These programs already contain a few integrated induction methods, which were primarily made for commercial web pages. Because of this the methods had to be redesigned in a way that they could be used for our e-learning environment. Most LMSs (Learning Management Systems) have an integrated module, with which albeit is possible to record specified elements of user activity such as data in connection with access, or the indicators of downloading particular elements of the curricula. However, these analysis methods are not sufficient for forecasting or revealing extensive coherences.

Data mining software products are primarily used for monitoring the usage of commercial web pages and forecasting user behaviour. So in the current business environment (however at a very high price), the integrated statistical elements of the software support the analysis of the log files of the web pages, and the formulation (and projection) of the results from a marketing point of view. The program's user interface and its usage do not necessarily require a qualification in IT, or in mathematics, but an interest in these subjects can be an advantage.

Fig. 1 shows the usage of some of the software's integrated nodes. From left to right these are the following:

– input database – which is Moodle's CSV³ stock for one (or more) given course, selected by us,

²INTegrated Evaluation and Research Facilities for Assessing Computer-users' Efficiency.

³A Comma separated values (CSV) file is a computer data file used for the digital storage of data structured in a table of lists form, where each associated item (member) in a group is in association with others also separated by the commas of its set. (http://en.wikipedia.org/wiki/Comma-separated_values)

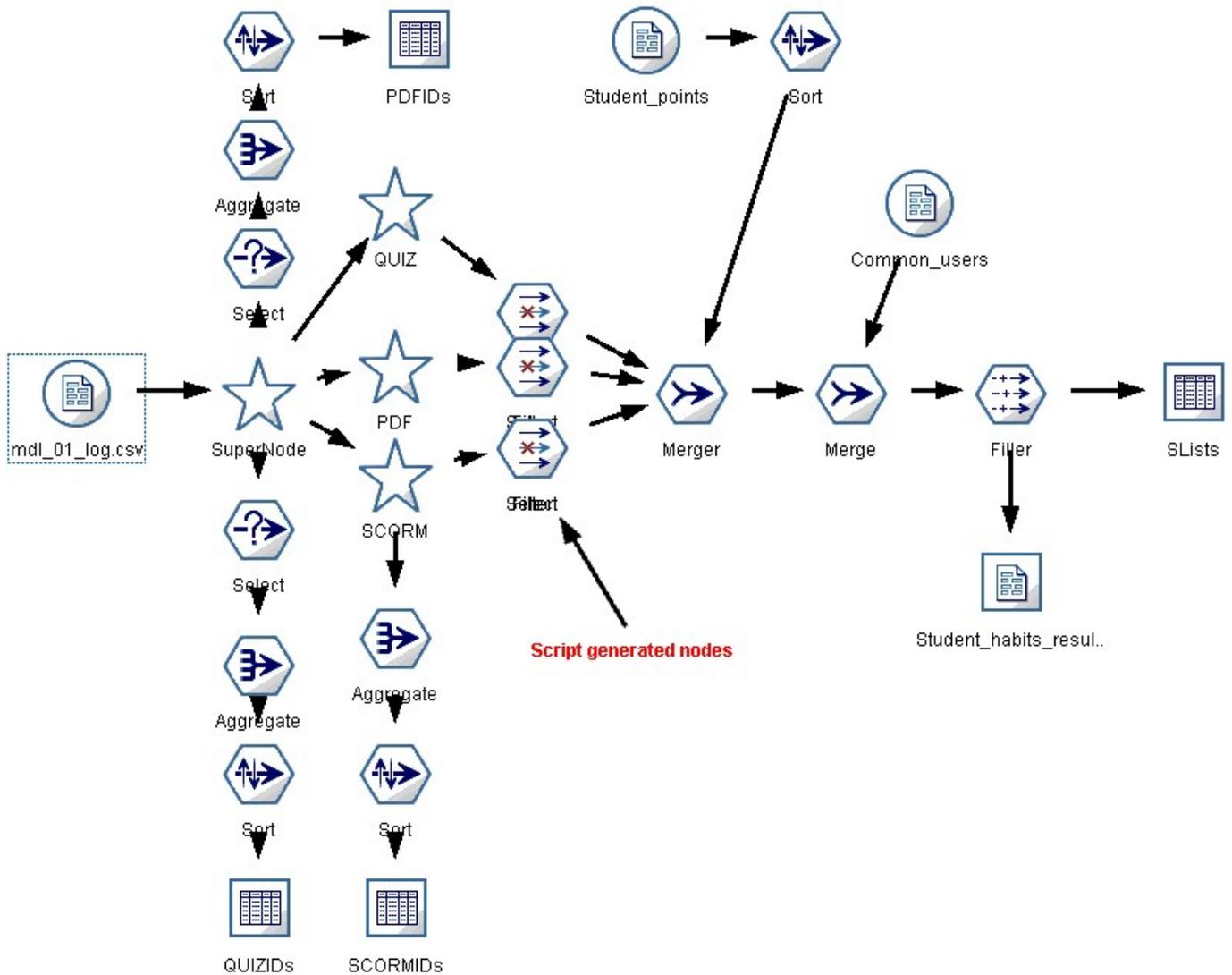


Fig. 1. Data analysis with the support of the SPSS Clementine 12.0 software

- a super-node, which by carrying out the default set of tasks, performs a data screening process in line with the given goals, whose parameters have been set.,
- the selected data can be subjected to further filtration – that are more goal-oriented,
- the aggregation of the data – connection can be established between the various data-panel’s rows and columns,
- data can be retrieved in a table form, which can be exported to further spreadsheet software (MS Excel, SPSS). Further nodes can be used for the graphic visualization of the results.

The practice and exchange of information with “fellow partners” surprisingly speed up the learning curve of the software. This is why a Web-Mining Research Group researching the software’s probable scope and development opportunities, operates at the Department of Ergonomics and Psychology, BME under the supervision of Dr. Lajos Izsó.

One of these scopes is the analysis of the usability of e-learning courses. The teachers and students are always asked to evaluate our distant learning and so called blended learning

type courses to get to know their opinions about this type of teaching, the curricula, and possible problems. The numerous criteria necessary for data mining are met – since:

- the user interface is very well known, because we have designed it,
- the users (students and teachers) are also very well known, and they can be identified if not personally, then by their NEP-TUN (university registration) codes,
- we do know what we would like to achieve by getting users to use the course interface, since we were the ones, who laid down the requirements for the course. Within the system we are also expecting tasks, and we are offering the opportunity of accessing curricula, and communication opportunities.

Our inexperience and the previously mentioned attributes of the software – namely that the data analyzing procedures were designed for commercial web pages – can present some difficulties. So they inevitably must be redesigned.

With the usage of data-mining methods the massive bulk of data can be transformed into expedient information, and

knowledge [3]. The Moodle electronic learning environment supports a number of spreadsheet systems, our three research groups mentioned in the Introduction use the MySQL spreadsheet. Moodle stores information in nearly 200 data-tables – so the expression „mining” is rather appropriate for defining this process.

The interpretation of the data, and the definition of the identified activities were carried out with the help of Clementine 12.0 data-mining software’s nodes (its intersections and its units containing a certain series of operations). In the following it will be demonstrated how the previously mentioned questions in regard with the use, and direct usability have been answered.

5 Questions

The questions mainly concern the determination of the quality of our educational portal and the contents within (course sheets, and elements of the curricula) – as a quality assurance consideration. Accordingly, the following quality factors have been defined:

- the usability factor, which can be defined by consistency, simplicity, perspicuity and software ergonomic aspects,
- effectiveness which is in connection with the efficiency of the participants’ work, learnability, the time spent, and meeting the pedagogic and personal needs,
- motivation, which is an ever occurring question in education – so cannot be left out of account. For the sake of increasing it or maintaining it communication between participants is strengthened.

In this study the results achieved with the help of the data-mining method concerning the effectiveness, the consumed time, and the usage of the communication channels will be presented.

6 The research

Moodle stores all the performed actions on its interface in a SQL⁴ database. This database records user identification and all the elements of the course-sheet, the usage and the used time span of them as well. The massive bulk of data (in some cases even a few Megabytes) contains the data in a „mashed” form. The important parts for the interpretation and the parts meeting the objective are to be distracted from this. This task is made easier by the application of the data-mining software.

So questions were formed that could not be answered by means of the integrated statistical tools of the supporting system, then (thus) we tried to answer these questions with the help of using data-mining software.

The Clementine software package has been used to answer the following questions:

⁴ SQL (Structured Query Language) is a database computer language designed for the retrieval and management of data. It is also a programming language for querying and modifying data and managing databases. (<http://en.wikipedia.org/wiki/SQL>)

- The detailed analysis of the activity of individual users (differentiated according to the roles) – the quantity and the distribution of all activities carried out on the course-sheet,
- The indicators of the available elements of the curricula – in addition to the number of the downloads, the analysis of the habits and tendencies,
- The indicators of the usage of complementary communication possibilities and other tools,
- The thorough examination of the phases of the course (independent learning, checking the knowledge, practical work),
- Identifying the students’ learning habits and comparing them to the results reached.

We had the opportunity of analysing the activities, trying to identify the learning habits of the 46 mutual students of our two courses (*Consumer Protection* course and *Planning for Special Users* course).

7 Answers

7.1 The comparison of the activities carried out on the course-sheet

The learning activity was examined by analysing the activities connected to the individual elements of the curricula. We were very curious to know the system activities (only that are connected to learning) of individual students during one visit. This obviously contains the usage of the professional forums, since if a student has a question regarding the interpretation of the curricula, he/she can pose the question to the tutor during studying. The examination of forum usage is outlined in the following subsections.

Table 1 and Fig. 2 show the common activity level of the students (students who attended both courses and tutors) enrolled in the two courses’ (subject of our analysis) and the points when these levels are outstandingly high or low.

7.2 The indicators of the tenure of the communication tools

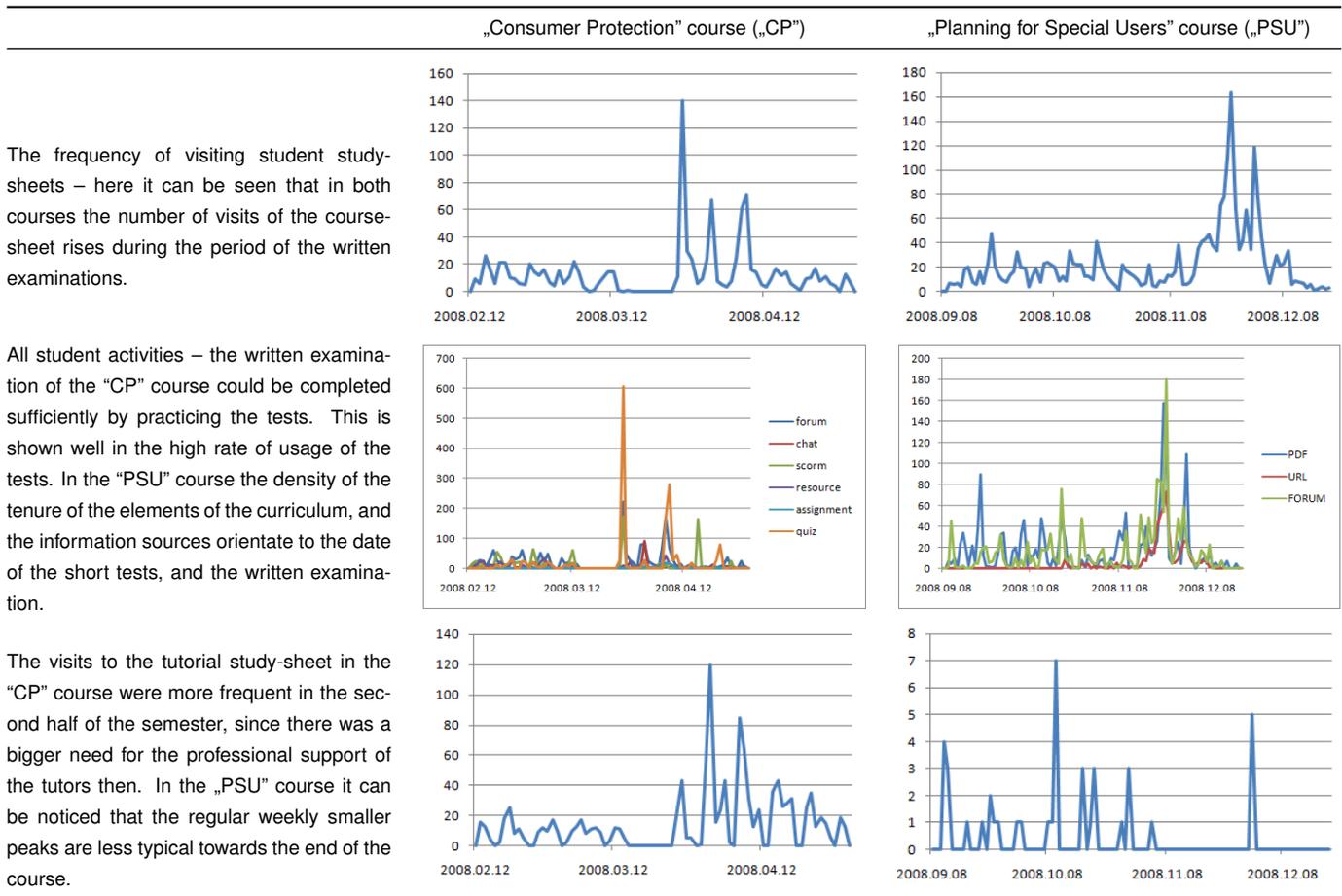
The electronic communication between the participants of the courses (students, course designers, organisers and tutors) has an important role, since apart from the first (introductory) lecture, no further personal contact takes place. For this very reason, a number of communication channels has been developed for exchanging, and sharing information. These are the following:

Forum – as a stage of professional questions and answers (it had a particular role in the “CP” course, but the formation of the groups, and the registration for the examinations were managed through the forum function as well).

Messages – a primary communication channel between tutors and students.

Chat – it was used as a consultation course once in the “CP” course.

Tab. 1. Tenure and activity of the course-sheets by subjects and roles



The frequency of visiting student study-sheets – here it can be seen that in both courses the number of visits of the course-sheet rises during the period of the written examinations.

All student activities – the written examination of the “CP” course could be completed sufficiently by practicing the tests. This is shown well in the high rate of usage of the tests. In the “PSU” course the density of the tenure of the elements of the curriculum, and the information sources orientate to the date of the short tests, and the written examination.

The visits to the tutorial study-sheet in the “CP” course were more frequent in the second half of the semester, since there was a bigger need for the professional support of the tutors then. In the „PSU” course it can be noticed that the regular weekly smaller peaks are less typical towards the end of the course.

Using the wiki discussion page – it was used during the preparation of individual assignments in the “PSU” course.

7.3 Analysing the use of the Forums

From the forums the professional (so-called “tutorial”) ones have been examined in the “CP” course. These forums served as a panel for the Q&A between the students and the tutors. Each special field had a separate forum.

During the first half of the course the students posed none too many professional questions, they were only “sampling” questions to the tutors (namely regarding the choice of the subject). During the second half of the course 3 peaks have been distinguished as shown in Fig. 2.

The first is the period of choosing the subject and preparing an outline. The second can be linked to the deadline of the revised outline, and the third shows the deadline of handing in the assignments.

7.4 Examination of the use of the Chat-room

The chat-function has been examined in the “CP” course. It was interesting to know how many of the students participated actively in using the functions supporting collaborative work, how many of the students “put a word in” for the individual subjects, and how many of them “kept themselves back” in the background. The chat-function was only an hour-long part of

a single e-tutorial during the course. As this function is so uncommon in the current conventional teaching practice, that it was thought more appropriate not to make its use regular in the course. From the 110 active participants of the course, a total of 69 participated in the chat, and only 12 students put a word in for the subject. However, one student responded 19 times to the chat.

The content of the chat was also available to the students after its shutdown – so they could read what the chat was about. The chat was visited 3 times on the day of the shutdown, and 4 times following it.

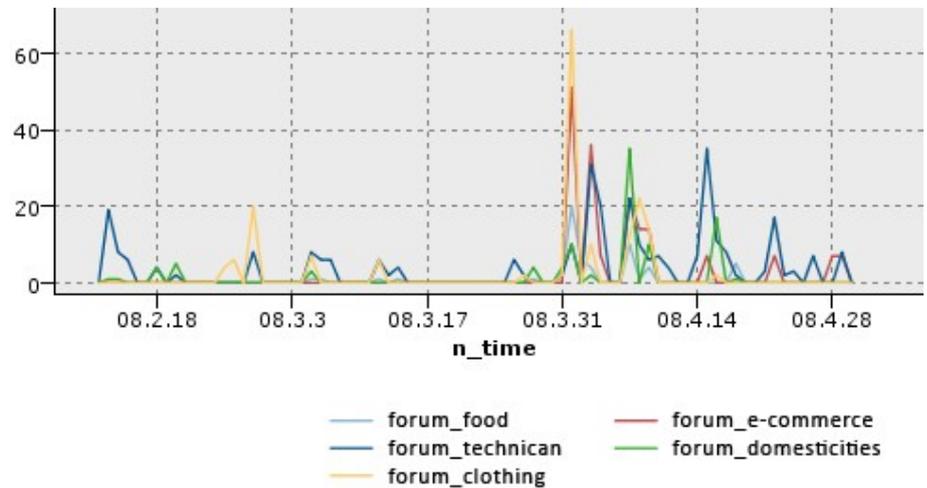
7.5 Analysis of the use of the Debate Page

The use of wiki elements supports the solving of the tasks collaboratively, in real time. The editors can keep track of who wrote, what and when to the task, and the tutor directing the task can also keep track of the happenings, and can make a remark, change, control or correct the mistakes.

At EPT’s courses the elements were extended by supporting learning by means of the MediaWiki function. However, the wiki function does exist in the Moodle as well, it is not as easy to use, as the above mentioned MediaWiki.

The MediaWiki – as well as the other mutual “information storages” – were filled with content by the students. By providing a collaborative work interface, our goal was to achieve that the students really get to know each other’s work, that the home

Fig. 2. Use of the professional forums



assignments are not only prepared for the teachers, but all the participants of the course (student, teacher, tutor) could benefit from them. The questions concerning the activity of the participants using the common storage of knowledge are the following:

- Who edited how much and when (in time; before, after, intensity)?
- Who went through other groups’ work?
- Who read the tutorial?
- Who used the discussion page (whether he/she only read it, or wrote as well)?
- How strongly is the frequency of the activities linked to the deadlines?

From these questions it is necessary to emphasize an important one: The wiki activities in the PSU course have been analysed. From the 46 mutual students of the two courses only 9 used the discussion page (wrote on it). So for the time being this form of communication is quite alien for the students. From the 9 students no one gave his/her opinion of the others’ work, their comments were related to the format of the final document, or it was in defence of their subject (parts with the same content could not be repeated).

During the performance of the individual assignment the use of MediaWiki was obligatory, while during the group assignment it was not. It is interesting to know how many of the students took a liking for completing this assignment with this interface. The results of the 6 groups are the following: two uploaded materials, but they came from the same person, and it was uploaded posterior – so the collective work did not go on through the interface.

8 Analysing the learning habits of the students

From the order of the learning activities the following conclusion can be drawn concerning the students’ learning strategy. For this it is important to observe that how often, and in what order the students use the individual curriculum elements (downloadable content, self-assessment tests, and contents that can be

viewed on the screen). By determining the frequency of usage of each curriculum element it can be found out what level of knowledge the students have reached. Can a typical order of the usage of curricula elements be detected? Can there be coherences detected between the order of use, and the results? If so, can the “successful” and “unsuccessful” learning strategies be identified?

For identifying the successful learning strategy it has been examined whether the frequency of the use of the curriculum elements influence the results at the end of the semester. Comparing the students’ (measurable) activity of the use of the curriculum, and their results at the end of the semester, it can be identified whether the activities labelled as a “successful” or “unsuccessful” strategy in the individual courses, are demonstrated in the results similarly in the two courses for a given student.

8.1 User activity concerning the results

During the association analysis rules that occur regularly among the variables (data-value) in a certain database have been sought [4, p. 201].

For analysing the successful learning strategy it has been examined whether the frequency of the use of curriculum elements, and the practice tests have an effect on the result at the end of the semester. (Acquiring the result did not happen within the frame of the system, but by completing the written examination and the practical assignment, where the practical use of the acquired knowledge was evaluated).

According to our assumptions, the more a student engages him/herself in the curriculum, the better result he/she will have. The outcome of the examination confirmed that the more tests the student completed, the better scores he/she had. But the same coherence was not confirmed between the sections of the curriculum and the acquired points. So this proves that the students were more successful at the written examination who completed more practice tests during the semester.

This assumption is only true in respect to the best students in accordance with these examination marks, but they did not necessarily acquire more in-depth knowledge.

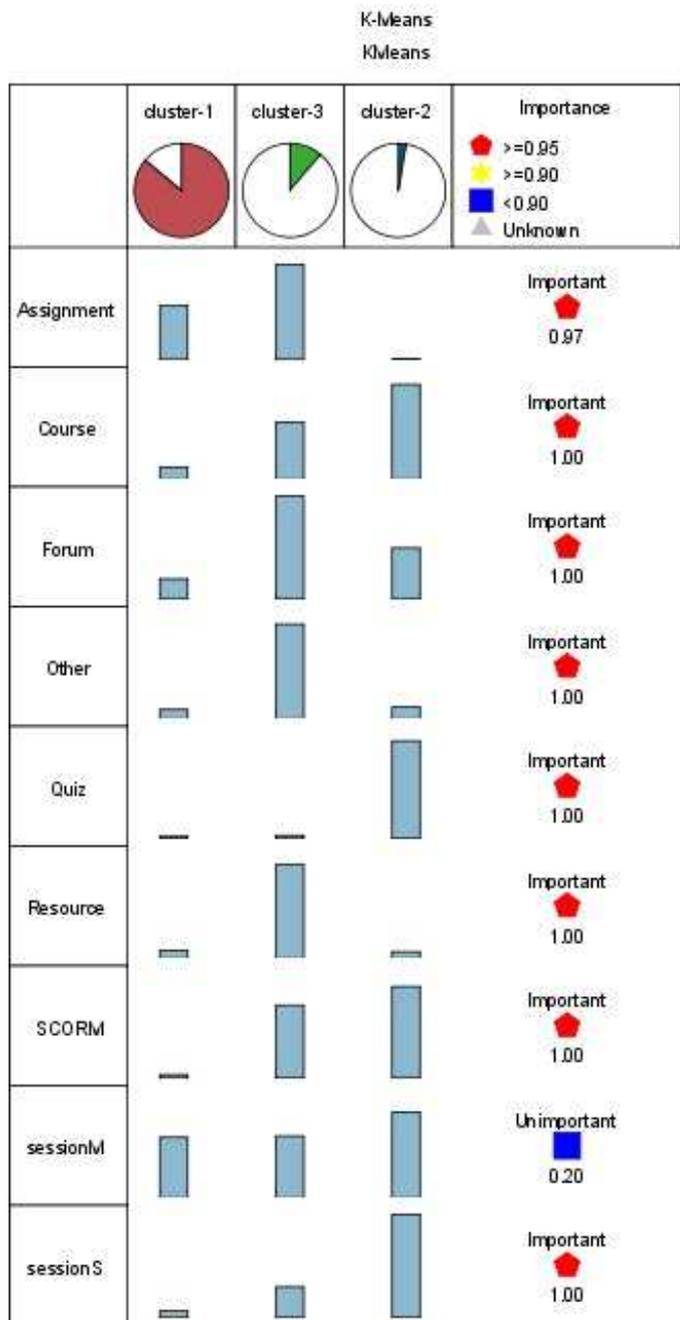


Fig. 4. The students of the CP course were categorised by functions

be differentiated.

It can be concluded that the participants of the first group (the largest by headcount) went through principally all of the curriculum elements, but none of these are of primary importance. In contradiction to our expectations these students did neither pay a lot of attention to the SCORM modules, nor to solving the tests. Their most important task during the semester proved to be implementing the registration for the exam, which was made possible on the subject's page. They can be called the „silent majority”.

In the activities of the second group, the elements of the curriculum weighed much more – except the practice tests. Apart from this, they could be the „dream group”, since they obviously used the opportunities of the Moodle electronic learning

environment, and curriculum edited by the teachers. They also relied on the tutorial help of the forums. Their alienation from the tests is a bit strange. Their name could be: „the group close to the top”.

The third and smallest group showed the most effective attitude. After getting to know the contents of the SCORM modules, they put their acquired knowledge to the use by completing the tests. They used the tutorial forums, and checked the data page of the subject relatively often, but the downloadable PDF contents left their interest unattracted. Their name can be: „effective small group”.

It is an interesting examination that while in this educational framework about 80% of the students try to „survive” the semester silently with the least effort, this rate is similar in the reality. Only the remaining 15-20% are willing to make efforts to meet the minimum requirements. It should be our (teachers') aim to change these rates.

9 Vision

Usability evaluation need not end with a system's release. Standard Web server, or httpd, logs are an invaluable source of information about the usage patterns once a Web site or learning environment has gone live. At this point the testers need not find usability experts or representative users; real users' sessions are captured in great details and are available for analysis.

Naturally every opportunity cannot be exploited in researching the in-depth coherences. By examining the order of tapping in learning objects of the curriculum, a general idea can be gained about the learning strategies of the students or the lack thereof. If a more plentiful source concerning the frequency of use of the curriculum elements is available, we can give a few more simple predictions as well (e.g. according to the activities carried out in the first few weeks, the students' learning styles can be identified, and personal advice can be given (rather concerning types and styles) in the hope of a successful outcome). We see managing the students as a community as a further possible opportunity; on behalf of this, certain socio-metric measurements can be made by using data-mining tools [4].

The use of the SPSS Text Mining for Clementine text-mining software would provide further possibilities of analysing the habits of communication. Although other tools (like Google Analytics, WebStat) provide cheap results of usage data, Clementine has the opportunity of re-building or modifying proof, operable streams according to special circumstances of our own courses. And also gives the opportunity to explore new or hidden coherences between course data.

The occurrences of issues, keywords, or notions in the chat-rooms or in the forums can imply which elements of the curriculum are hard to interpret, or are interesting for the students. The plug-ins (extensions) compatible with the Moodle learning environment make it possible to support the collective learning processes with elements missing from the Moodle, or elements available that are more sufficient to use. That is why, we de-

cided to integrate the MediaWiki to the educational procedures – which in our opinion was very successful amongst students. But our plans include a forum supported with a web-cam, the possibility of establishing a web-conference, if there is a relatively low number of participants (for example a group of ten) sits down one time to study or do an activity. This solution has the disadvantage of only having normal quality visuals in the case of a broadband Internet connection. Instead of the currently used chat function, a “Moodle chat” supplement could be used, which would (as most of the chatrooms) support real time communication and file sharing between the participants. A number of similar applications have been tested, in the near future the application of the Skype-module will be introduced.

The widespread usage of electronic learning environments has a significant affect on learning and teaching activities these days. The permanently increasing number of ICT tools changes educational technology and applications almost day after day. Besides this, learning habits and strategies are also changing. Analyzing the data of log files is the easiest way of identifying and deducing the students’ learning habits in an e-learning-based environment. Web usage mining techniques have been used to explore the coherence between usage data and usability in order to find opportunities for development that increase the efficiency in e-learning environments. I would like to keep on "mining" to collect enough analyzing models and results that can properly complete traditional usability testing methods and compose the requirements of an effective learning environment and user interface.

10 Conclusion

Our theory was that results of analyzing data log files of electronic learning environment, the usage of the content (curriculum) and the user interface could help us in creating an effective system for independent learning process. We suppose that we can establish the learning environment that is meeting personal requirements by collecting data logs of individual learning activities and results. We can compare individual activities with our expected learning activities. Our expectations mean that in the interest of a successful execution the student have to complete certain activities (readings, tests, forums, etc.) On the other hand, we have to consider the students’ own learning styles and habits in order to offer them personalized learning instructions.

Analyzing method we have presented in this study used the data of users’ activities and the system (user interface and learning content) parameters. Among these data we would like to identify those ones which could influence personal performance. We suppose we can use these data for creating an assistive system for independent learning in electronic learning environment.

References

- 1 **Izsó L, Antalovits M**, *Questions of quality assurance of educational software products*, Introduction to information-ergonomics, BME, 2000, 171–173.
- 2 **Levi M D, Conrad F G**, *Usability testing of World Wide Web sites*, (2008), http://stats.bls.gov/ore/htm_papers/st960150.htm.
- 3 **Hanne M**, *Data mining in the e-learning domain*, Campus-Wide Information Systems, **21**, (2004), 29–34.
- 4 **Dunham N H**, *Data Mining: Introductory and Advanced Topics*, Prentice Hall, Upper Saddle River, NJ, 2003.
- 5 **Jókai E, Horváth J C, Nagy G Z**, *Analysing the learning habits by data-mining methods*, APPI Conference (November 2008).