VILNIAUS UNIVERSITETAS

Saulius Preidys

DUOMENŲ TYRYBOS METODŲ TAIKYMAS SUASMENINTO ELEKTRONINIO MOKYMO APLINKOSE

Daktaro disertacijos santrauka

Technologijos mokslai, Informatikos inžinerija (07 T)

Vilnius

Disertacija rengta 2008–2012 metais Vilniaus universiteto Matematikos ir informatikos institute.

Mokslinis vadovas

prof. habil. dr. Leonidas Sakalauskas (Vilniaus universitetas, technologijos mokslai, informatikos inžinerija – 07 T).

Disertacija ginama Vilniaus universiteto Matematikos ir informatikos instituto Informatikos inžinerijos mokslo krypties taryboje:

Pirmininkas

prof. habil. dr. Gintautas Dzemyda (Vilniaus universitetas, technologijos mokslai, informatikos inžinerija – 07 T).

Nariai:

prof. dr. Eduardas Bareiša (Kauno technologijos universitetas, technologijos mokslai, informatikos inžinerija – 07 T),

dr. Vaida Bartkutė-Norkūnienė (Utenos kolegija, fiziniai mokslai, informatika – 09 P),

prof. dr. Valentina Dagienė (Vilniaus universitetas, technologijos mokslai, informatikos inžinerija – 07 T),

doc. dr. Daiva Vitkutė-Adžgauskienė (Vytauto Didžiojo universitetas, fiziniai mokslai, informatika – 09 P).

Oponentai:

doc. dr. Vitalijus Denisovas (Klaipėdos universitetas, fiziniai mokslai, informatika – 09 P),

prof. dr. Dalius Navakauskas (Vilniaus Gedimino technikos universitetas, technologijos mokslai, informatikos inžinerija – 07 T).

Disertacija bus ginama viešame Vilniaus universiteto Informatikos inžinerijos mokslo krypties tarybos posėdyje 2012 m. gruodžio 19 d. 13 val. Vilniaus universiteto Matematikos ir informatikos instituto 203 auditorijoje.

Adresas: Akademijos g. 4, LT-08663, Vilnius, Lietuva.

Disertacijos santrauka išsiuntinėta 2012 m. lapkričio 19 d.

Disertaciją galima peržiūrėti Vilniaus Universiteto bibliotekoje.

VILNIUS UNIVERSITY

Saulius Preidys

THE APPLICATION OF DATAMINING METHODS TO PERSONALISED LEARNING ENVIRONMENTS

Summary of Doctoral Dissertation

Technological Sciences, Informatics Engineering (07 T)

Vilnius, 2012

Doctoral dissertation was prepared at Institute of Mathematics and Informatics of Vilnius University in 2008-2012.

Scientific Supervisors

Prof. Dr. Habil. Leonidas Sakalauskas (Vilnius University, Technological Sciences, Informatics Engineering – 07 T).

The dissertation will be defended at the Council of the Scientific Field of Informatics Engineering at the Institute of Mathematics and Informatics of Vilnius University:

Chairman

Prof. Dr. Habil. Gintautas Dzemyda (Vilnius University, Technological Sciences, Informatics Engineering – 07 T).

Members:

Prof. Dr. Eduardas Bareiša (Kaunas University of Technology, Technological Sciences, Informatics Engineering – 07 T),

Dr. Vaida Bartkutė-Norkūnienė (Utena University of Applied Sciences, Physical Sciences, Informatics – 09 P),

Prof. Dr. Valentina Dagienė (Vilnius University, Technological Sciences, Informatics Engineering – 07 T),

Assoc. Prof. Dr. Daiva Vitkutė-Adžgauskienė (Vytautas Magnus University, Physical Sciences, Informatics – 09 P).

Opponents:

Assoc. Prof. Dr. Vitalijus Denisovas (Klaipėda University, Physical Sciences, Informatics – 09 P),

Prof. Dr. Dalius Navakauskas (Vilnius Gediminas Technical University, Technological Sciences, Informatics Engineering – 07 T).

The dissertation will be defended at the public meeting of the Council of the Scientific Field of Informatics Engineering in the auditorium number 203 at the Institute

of Mathematics and Informatics of Vilnius University, at 1 p.m. on 19th of December 2012.

Address: Akademijos st. 4, LT-08663 Vilnius, Lithuania.

The summary of the doctoral dissertation was distributed on 19th of November 2012.

A copy of the doctoral dissertation is available for review at the Library of Vilnius University.

1. Introduction

1.1. Research field and relevance of the problem

Contemporary information technologies are applied in various life areas: business, media, daily life, science, education, etc. Already in the high schools' pupils use personal computers and smart devices for collecting information, preparing presentations, communication. Not so long ago the methods of the second generation web (WEB 2.0) such as WIKI, BLOG, social networks were widely used but nowadays the era of the third web generation – data era (WEB 3.0) has started: systems become smart ones, internet users receive personalised and needs-adapted information. Systems are observing and analysing user's activities and suggesting appropriate ways of behaviour in case of necessity. This is widely used in smart search systems (Yahoo!, Google, Hakia), semantic social networks (Facebook, GroupMe!, Twins) and e-business systems.

Life-long learning become a necessity while acceleration of daily life tempo and increasing of information flows. Distance education becomes popular all over the world with software and Virtual Learning Environments (VLE) at use, which allow to study in the distance manner. While using VLE, the user can study in appropriate time and tempo and user becomes independent from the studies' location.

Widely used VLEs such as Blackboard Vista, Moodle, Doceos and others give a possibility for the learners not only to study the materials given but also take the self-evaluation tests, participate in the discussions and forums, fill in blogs and publish their observations in WIKI, etc. Another useful VLE function is observation of the learner's activities. Every learner's activity is fixed in VLE databases and/or journals. Course teachers and tutors can find the statistics of all users' activities analyse their activities and make changes in the study process in case of necessity.

The observation tool is integrated into almost all VLEs. In spite of that those tools reveal much less information, as it is accumulated in the VLE. Administrators of VLEs often simply delete this data from servers thus loosing a lot of useful information. It is possible to analyse data using various statistic methods, but as amounts of data become really big, *Datamining* methods are to be used, such as clusterisation, classification, visualisation, neural networks, etc.

In this research the tasks of the learners' activities' in VLEs data preparation for the datamining methods, integration of datamining methods into the open source VLE Moodle and the visualisation if this data is investigated. Also the new data visualisation area – *learnograms* are analysed.

Before planning to prepare and submit distance education course the initiators should pay attention to the fact that students are studying using various methods: some of them start reading the given materials one by one, the others look for the unfamiliar materials, the others go for virtual discussions, etc. Therefore, after analysing learning activities and identifying the learner's style it is possible to prepare personalised study materials and choose better ways for course presentation. Such learning organisation would improve the quality of studies and allow achieving better results.

Learning styles establishment is a very important element of educology seeking higher study outcomes. Two methods are used to identify the learning styles: communication method and automatic method. Using the first method the learners have to answer the questionnaire for the learning styles identification. After processing the data obtained, the teacher assigns each learner to one or another learning style. Using the second method, the self-evaluation questionnaire is skipped. That allows to evaluate the learners in a more objective way as possible incorrect answers are eliminated. Using the automatic learning style identification method, the learners' activities are analysed, information about the learners' behaviour is accumulated and similarities between the learners are looked for. The learners' activities records in VLE are very useful for the automatic learning style identification.

In this dissertation the following main problems are analysed:

1. Learners' activities in VLE data preparation for the datamining methods application.
2. Learners' activities data visualisation and visual presentation for the VLE users: administrators, teachers, tutors and students.
3. Learners' learning style identification analysing the accumulated data on learners' activities in VLE.

1.2. *The aim and objectives of the research*

The overall aim of this research is to create a methodology for the datamining methods' application in personalised virtual learning environments.

To achieve the aim, the following objectives have been set:

1. Research on distance education warehouses preparing application of the datamining methods.
2. Learners' activities evaluation by using learnograms.
3. Learners' learning styles identification.
4. Creation and application of personalisation methods of teaching.
5. Creation of programming prototype and integration of separate modules to the open source VLE Moodle.

1.3. The research object and methods

The dissertation research object is application of datamining methods to the data of learners' activities in the VLEs. This object is relevant for the intellectual data analysis or datamining as after interpreting the results achieved it is possible to prepare models and on the basis of the models improve the quality of distance education.

The main research methods applied in the dissertation are: observation, documents' analysis and experiment. During the research, the real data from VLEs' of Vilnius College and Vilnius University was taken and students, as well as their activities during the lectures while using distance education methods, were observed.

1.4. Scientific novelty

1. Systemic analysis of application of datamining methods to the data of learners' activities accumulated in VLE.
2. A method of VLE users' activities clasterisation, applicable for analysing the learning activities, worked out.
3. A rarely used method of learnograms, used for visualisation of VLE users' activities, which allows to observe all learner's activities and achievements during the selected period.

1.5. Practical significance of the work

The research results of this dissertation will be used while integrating computer agents into VLE Moodle: according to one of the activities of the Lithuanian Distance Education Consortium (LieDM) project „Lithuanian Distance Education Consortium network development", programmed modules will be integrated into VLE Moodle and will be used by all LieDM Consortium members, 14 Lithuanian universities, colleges and other education institutions.

Applying the automatic questionnaire method and using Case-Based Reasoning and Honney and Mumford earning styles classification, the module was prepared and

integrated into VLE Moodle, which allows to identify learner's learning style. Based on this method the program agent was prepared, which gives suggestions to the course tutor.

1.6. Approval of the results

Research results have been published in 6 scientific publications. 1 article has been published in foreign scholarly publication included into the Institute's of Scientific Information list of main journals with the citation index, 5 articles have been published in reviewed Lithuanian and foreign publications.

The author has participated and presented his research results in 11 national and international scientific conferences: Baltic DB&IS 2012 (Tenth International Baltic Conference on Databases and Information Systems), 5th scientific-practical conference „The Impact of Applied Scientific Research to the Quality of Contemporary Studies" (conference organised by Lithuanian Computer Society), an international conference „E-Education: Science, Study and Business 2010", LOTD-2010 (the 3rd Lithuanian Society of Young Researchers conference "*Operations Research for Business and Social Processes LOTD – 2010*"), an international conference „*Innovation and creativity in e-learning*", KODI-2009 (conference organised by Lithuanian Computer Society) and others.

The complete list of articles and conference proceedings on the name of Vilnius university, Institute of Mathematics and Informatics, is presented in dissertation's approval of the results.

1.7. Structure of the dissertation

Dissertation consists of the introduction, research area description, six chapters dedicated to the presentation of the research results, conclusions and the list of referred resources.

In the first dissertation chapter the problem of application of datamining methods in distance education and its relevance is presented the aim and the objectives of the research are formulated. In the second chapter of the dissertation, contemporary distance education architectures are presented, as well as hardware and software, used for those studies and the stakeholders of this process. In the third chapter, the datamining methods are presented together with their application in various areas. More attention is paid to the application of datamining methods in distance education; application stages and

possibilities are reviewed. In the fourth chapter the new data visualisation method in distance education – learnograms, used for the learners' activities analysis is presented together with its' practical application. In the fifth chapter learners' learning styles analysis is described together with the datamining application for identification of the automatic learning style. In the sixth chapter the workable prototype of students' activities analysis is presented, as well as practically described module of data analysis, which will be integrated into VLE Moodle.

2. Architecture of Distance Learning Systems

2.1. Hardware and software of distance studies

The following definition of Virtual Learning Environment is presented in the literature: *Virtual Learning Environment (VLE) – the education systems based on computer networks and other information technologies, where learners' study with the supervision of the tutors.* This is an environment where one can reach materials, take the tests and various assignments in the suitable time and tempo.

VLE is suitable for various projects' activities in-service courses, as well as additional resource for the strengthening of the knowledge and skills.

There are intellectual virtual learning environments examined in the dissertation. There are no full intellectual learning environments realised yet, but some have already been created and used. Those are so called *intelligent tutoring systems*. Intelligent tutoring systems are such systems where the learning process is managed by the system itself, not by the user. That is to say that system themselves decide which materials, manner or method of learning should be presented for the user. The following systems can adapt to the concrete learner's needs and take into account the learner's prior knowledge.

In case when the database of knowledge contains the data about the lecture materials in various formats (texts, formulas, theorems, a broad info materials, etc.), the following learning environment (or intellectual advisor) could present the learning materials for the learner, in the structure as it is indicated in the picture 2.1.
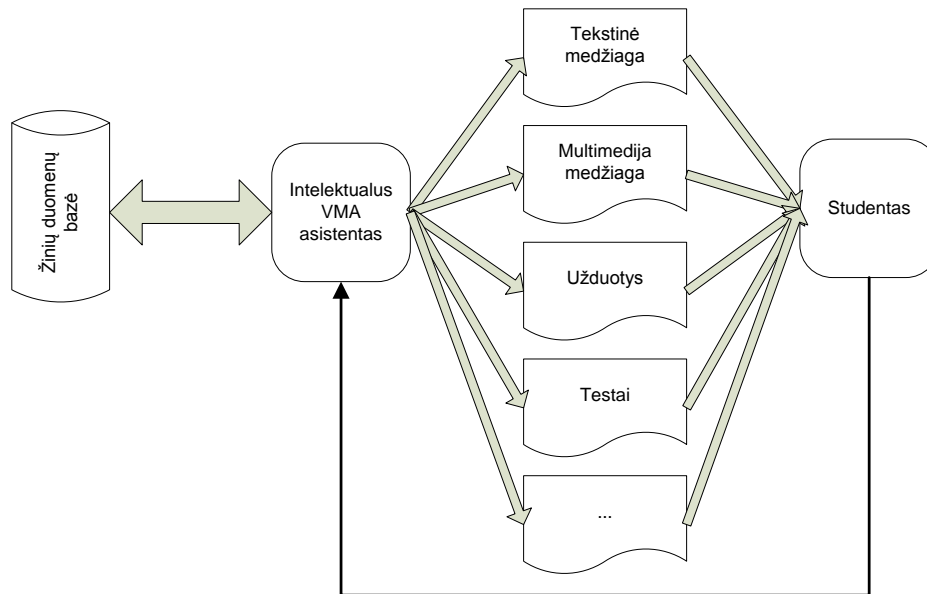
Picture 2.1. Scheme of cooperation between the intellectual assistant and VLE user

It is important to note that in such VLEs there is a possibility for the learner to choose a learning process according to his/her skills and perception or learning style. In the other learning environments such possibility is realised automatically, i.e., based on the features described above the system chooses the optimal learning process itself.

2.2. Users distance studies and their needs

There are different users participating in the distance studies process: learners, teachers, authors of distance modules, auditors, VLE administrators. The quality of distance studies depends on all of those groups participating in the study process. However, as the number of distance course supply and distance learning students is constantly increasing it becomes more and more difficult to observe the study process. For this reason, the quality of distance studies is declining, and the students are loosing their interest in the distance education; the quality of the tutors' performance is declining, as well.

The author of this dissertation is using his own experience and experiences of many scientists working in the distance education area and using various VLEs and various users groups: learners, tutors, course authors, auditors, VLE administrators, guests. Every user group performs their own activities in the study process. In the dissertation, all those distance education users groups are analysed in details, their needs are identified and the ways of their problems' solutions are suggested.

11

2.3. Data warehousing of distance education

It is necessary to receive and prepare VLE data before applying datamining methods to them. However, a big problem arises, as data in various VLEs is preserved in different ways. Willing to apply datamining methods to the available VLE data, there is not enough to use one table. Usually several data tables are needed. Therefore the data warehousing are suggested to be used for VLE data accumulation.

A. Data warehousing

Nowadays many companies meet the problem of information overload – there is a lot of information, usually too much, but it is not harmonised or systemised; not always reliable and not quickly achievable. The same situation applies to the VLE information. For the solution of the problem mentioned above the *Data Warehousing* concept is being suggested.

B. The architecture of data warehousing of distance education

The architecture of data warehousing of distance education slightly differs from the traditional data warehousing. The data warehousing used for the research of this dissertation is presented in the picture 2.2.
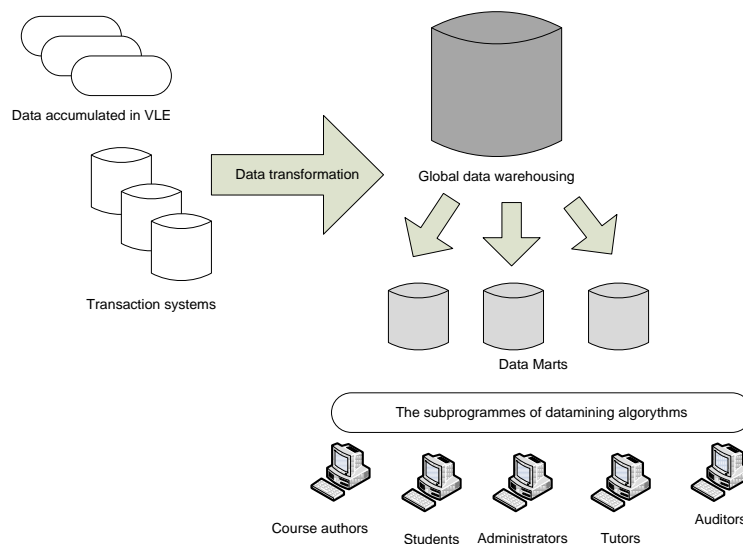
Picture 2.2. The scheme of Data Warehousing

As it was mentioned above, the data marts are to be used for implementation of the data warehousing. This method was chosen because of the convenience of information presentation to different users.

Data transformation and decoding is also being implemented during the data transfer from VLE to the data warehousing. In the global data warehousing the data is allocated into separate data marts: some tables contain data for the users behaviour identification, the others – for the learners' styles identification, the third – for the learning achievements analysis, etc.

The advantage of the following system is that it is universal and can function with different VLEs. It is only needed to create and adapt the transaction system to the environment for the system in order to function with the other VLEs.

Intensive usage of VLE and a big number of teachers, courses and students in the VLE server create a big amount of data in Moodle database. For example, in June 2012 in VLE of Vilnius University there were around 800 courses, 20000 users and 6 mln users' activities' recordings accumulated during one academic year.

3. Integration of datamining methods in the distance studies

The concept of *Datamining* appeared in 1978 and it acquired great popularity during the first decades of 1990. Datamining is a process of using various data analysis tools, which helping to discover data structures and relations, which could help in defining real conclusions and results.

The main datamining processes:

1. Classification;
2. Evaluation;
3. Prognosis;
4. Grouping according to the common features or relations rules;
5. Clustering;
6. Description and visualisation;
7. Neural networks.

Data warehousing technology is used successfully in business, medicine and other areas where big data amounts are to be processed. Several years ago datamining methods are starting to be used in educology, as well. This became especially actual when distance education became a popular study method, and VLEs started to be widely used in the study process.
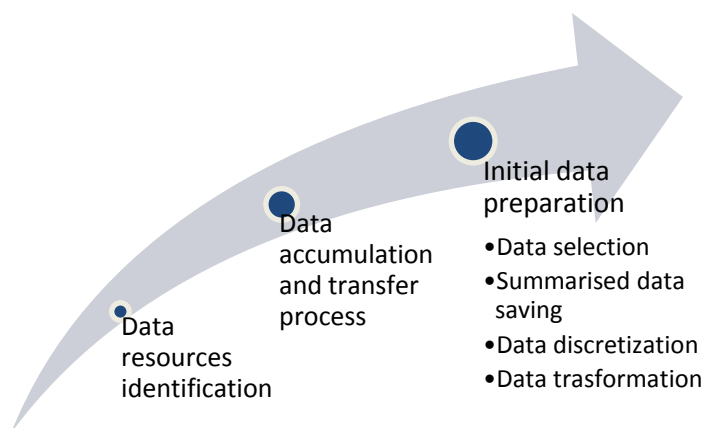
3.1. Datamining stages in the distance education

In the dissertation the datamining application in the distance education is analysed, which consists of four stages:

1. **Data accumulation.** VLE Moodle fix students activities in the activities journal files and MySQL, Oracle or other databases.
2. **Preparation of the accumulated data.** Accumulated data is cleaned and transformed to the appropriate format, which is understandable for the datamining software.
3. **Datamining methods application for the data obtained.** Applying datamining methods a lot of useful information is obtained for all the VLEs users.
4. **Interpretation and analysis of the data obtained.** Teachers and tutors can adjust the study process after analysing the data obtained, as well as change teaching methods and achieve better results.

3.2. Data accumulated in the VLE preparation for the datamining

Data accumulated in VLE preparation for datamining is a very important stage of datamining application for the VLE data. Data preparation scheme is represented in the picture 3.1.

Initial data preparation
- Data selection
- Summarised data saving
- Data discretization
- Data trasformation

Data accumulation and transfer process

Data resources identification

Picture 3.1. Data preparation scheme

3.3. Data source identification

Data sources in VLEs are a little bit different from those of other WEB systems. Most of the systems users' activities data accumulate in the activities journals files. Whereas VLEs all this info accumulate not only in activities journal files but in their own databases also. Therefore, VLE accumulate more info about their users' activities than other WEB systems. Those systems can save information about when and what was read or written by a student, when he/she took tests, how much time was spent on that,

how much time was spent on each question was the question repeatedly answered, where there any trials to discuss in the forums, synchronic discussions, etc.

The main data source in VLE Moodle is users' activities accumulation table **mdl_log**. The following users information is accumulated in this table (picture 3.2.):

1. Login time;
2. User identification;
3. Computer IP address;
4. Moodle course number;
5. Module, which was accessed;
6. CMID, course module Id;
7. Activity, which was performed during the current login;
8. Address, which was used for the login;
9. Additional information (test name, course number, etc.).

id	time	userid	ip	course	module	cmid	action	url	info
1	1267098039	2	127.0.0.1	1	user	0	update	view.php?id=2&course=1	
2	1267098081	2	127.0.0.1	1	course	0	view	view.php?id=1	1
3	1267098119	2	127.0.0.1	1	course	0	view	view.php?id=1	1
4	1267098151	2	127.0.0.1	1	course	0	new	view.php?id=2	testu analize (ID 2)
5	1267098158	2	127.0.0.1	2	role	0	assign	admin/roles/assign.php?contextid=10&roleid=1	Administrator
6	1267098167	2	127.0.0.1	2	role	0	assign	admin/roles/assign.php?contextid=10&roleid=2	Course creator
7	1267098177	2	127.0.0.1	2	role	0	assign	admin/roles/assign.php?contextid=10&roleid=3	Teacher
8	1267098182	2	127.0.0.1	2	role	0	assign	admin/roles/assign.php?contextid=10&roleid=5	Student
9	1267098189	2	127.0.0.1	2	course	0	view	view.php?id=2	2

Picture 3.2. Users' activities accumulation table *mdl_log*

Nevertheless, this only table is not enough for deeper analysis of users' activities. The other users information, such as evaluation, participation in synchronic and a synchronic discussions, data on the assignments performance, participation in the group work, etc., is accumulated in other tables. VLE Moodle contains over 300 different relational tables.

3.4. Datamining data preparation stages

In case of willingness to apply datamining methods for the data accumulated in VLE Moodle, a special data preparation is necessary. At this stage necessary data is selected and a table is formed, where aggregated data is presented. Some numerical data is discretised and transformed at the final stage.

Data selection. Data selection for the further datamining methods application depends on the research and activities to be performed.

All teaching objects usually are selected but there are users, whose activities can distort the statistics. Those are the most active and the most passive users. Therefore, the following exclusions will be eliminated or analysed separately.

Elimination of the outliers. The elimination of the outliers has a very important role for the VLE users' exact data analysis. As VLE users can be very different, their activities strongly differ, too: Some of users participate in various activities very actively, while the others login for the first time only before testing. Some of VLE objects are not used in study process at all. Therefore, wrong results can be obtained if outliers are not taken into account.

It is notable that sometimes those outliers give a lot of information, therefore it is not possible to ignore them. For example, using clusterisation algorithms for the data accumulated, outliers usually create separate clusters, therefore the analysis of those clusters is necessary.

For fixing of those activities, the data of various measurements is kept, like evaluation, login time and frequency, access to learning objects', etc. Standardisation is used for the data unification. The most widely used data standardisation is calculation of z-values. Having a set of data $x_1, x_2, ..., x_n$, z-values are calculated by using the formula:

$$z_i = \frac{x_i - \bar{x}}{S}, i = 1, 2, ..., n \qquad (1)$$

here: $\bar{x}$ – arithmetical average; S – standard deviation.

In this case, tolerance can be value, which is derived from the sample in two or three standard deviations. An observation x_i is an outlier in case of $z_i > 3$.

Preservation of the summarised data in the separate table. A separate table is necessary to preserve summarised data for the operations needed. For example, for the common students' activities and behaviour in VLE identification only such data is selected, which directly shows the following elements: user's login time to VLE, duration, access to the concrete learning object (test, learning materials, self-evaluation tests, discussions, etc.).

In the table 3.1. the summarised data is presented which will be used for student's activities analysis.

Table 3.1. Summarised user's data

No	Feature	Value
1.	Student	Student's unique code
2.	Assignment	Work with the assignments given
3.	Blog	Using of BLOGs
4.	Course	Number of students' logins to VLE
5.	Forum	Virtual discussions tool
6.	Glossary	Learning aims and objectives
7.	Questionnaire	Participation in teacher's questionnaires
8.	Quiz	Tests and self-evaluation test
9.	Resource	Work with the learning resources given
10.	Upload	Upload of the given tasks' solutions
11.	User	Review of student's activities in VLE
12.	Mark	Student's evaluation

Data discretisation. This activity is used for data clearness and concreteness. Discretisation method is distribution of available data to separate categories. Categorical variables are much more concrete for every user. Evaluation's distribution to the categories can serve as an example of discretisation: if the evaluation is < 5, the value assigned is FAILED, if the evaluation is between 5 and 7, the value assigned is GOOD, etc. the other categories can be used for other fields, e.g. for student's activity the following categories can be used: low, average, high.

Data transformation. Before applying datamining methods to the data available, data is transformed to the format suitable for the performance of the appropriate algorithm.

Data decoding can serve as an example of data transformation. For example, data decoding can be performed for the reduction of primary data diversity merging this data. It is possible to decode the current variables as well as to create new variables on the basis of the decoded values.

In this research the following decoding ways are used: decoding values without changing the variable and decoding values with the introduction of new variables. This allows to present the values in a more optimal way or to merge them.

3.5. Datamining methods in the distance education

There was a lot different research performed during the dissertation writing process and the various datamining methods were applied to the VLE users' activities data accumulated.

Further the following datamining methods will be discussed: clusterisation and multidimensional association rules.

3.5.1. Clusterisation

Before planning to prepare and submit the distance education course, a course developers should pay attention to the fact that students are studying in different ways: some of them start reading materials from the beginning; the others only pay attention to the unclear topics; the third are using virtual discussions etc. Therefore, after analysing learning activities, later on it is possible to present him/her personalised learning materials, to choose more proper methods for course delivery. Such learning organisation would improve the quality of studies and would let to achieve higher results. Application of the clusterisation *k-mean* algorithm is analysed in this research using learners' behaviour in the most popular VLE Blackboard Vista in 2008-2010 in Lithuanian universities and colleges. This VLE was selected because of the existence of real distance education courses at Vilnius Gediminas Technical and Kaunas Technological Universities. Blended courses cannot present clear information about the learners' activities in VLE as part of the lectures is taught in classrooms in the traditional way.

During the research, the activities and results of 528 students in 15 distance education courses were presented. After applying k-means clusterisation, three VLE learners' categories were identified and the appropriate methodologies were suggested for those students' categories. In addition, the key factors influencing students' final evaluation were clarified.

A. Application of k-means clusterisation method to the VLE users groups identification

The most simple and widely used k-means algorithm is used for the data clusterisation. This algorithm is using formula (2) for the data clusterisation. Using this algorithm objective function O is minimised.

$$O = \sum_{j=1}^{k} \sum_{i=1}^{n} \left\| x_i^{(j)} - \mu_j \right\|^2 \qquad (2)$$

here: $\left\| x_i^{(j)} - \mu_j \right\|^2$ is Euclidean distance between the point $x_i^{(j)}$ cluster centre μ_j.

While applying k-means algorithm VLE users' activities are collected, as well as calculations of how many times the user accessed the system and used any of VLE tools are performed. VLE tools are analysed and presented in the table 3.2. In the following stage a vector matrix is formulated from the data calculated (Picture 3.3.).

B. Preparation of the accumulated data

After analysing the available data, observation of those VLE tools was performed, which indicated peculiarities of learners' activities. A part of the tools were rejected because of the low usage. In the table 3.2. the selected learners' activities attributes are indicated, which later on were used to identify the learners' learning style and activities.

Table 3.2. VLE tools used in the research

No	Attribute	Mean
1.	FINAL_GRADE	Final student's evaluation in the course
2.	CON_COUNT	Number of students' logins to VLE
3.	TIME	The time spent in VLE
4.	ANNOUNCEMENT	Information announced to the student
5.	ASSESSMENT	Tests and self-evaluation tests
6.	ASSIGNMENTS	Assignments
7.	CALENDAR	Calendar
8.	CHAT	Chat
9.	CONTENT_PAGE	Work with the materials given
10.	DISCUSSION	Tool of virtual discussions
11.	FILE_MANAGER	Tool of data files
12.	LEARNING_OBJECTIVES	Learning aims and objectives
13.	MAIL	Mail
14.	MEDIA_LIBRARY	Additional materials in video, HTM formats and dictionaries
15.	MY_GRADES	Grades
16.	MY_WEBCT	Tool of systemic settings
17.	NOTES	Reminders published for the students
18.	ORGANIZER	Learning module
19.	SYLLABUS	A plan of module topics
20.	STUDENT_BOOKMARKS	Students bookmarks
21.	TRACKING	Tracking student's activities in VLE
22.	WEB_LINKS	Additional resources – links to external web resources
23.	WHO_IS_ONLINE	Tool indicating current online module users

In the picture 3.3. the frequency of VLE tools usage is presented:

1 final_grade	2 conn_count	3 time	4 content_page	5 assessment	6 organizer	7 assignments	8 calendar	9 learning_objectives	10 discussion
1	5	10041	0	26	47	2	0	9	2
2	8	22441	0	51	81	4	0	2	1
2	63	73557	265	804	771	20	6	5	16
2	2	3433	0	7	4	0	0	0	0
3	14	9927	70	12	44	7	0	0	2
3	15	35746	46	11	151	98	11	0	94
3	4	3760	0	6	20	6	0	5	2
4	50	67480	115	38	174	79	37	7	98
5	12	12900	18	9	52	9	0	0	0

Picture 3.3. Data prepared after using clusterisation algorithm

C. Datamining methods application to the data received

Datamining and statistical analysis software STATISTICA was used for the data clusterisation. All VLE Moodle objects are used for the research. The aim of this research was to establish the influence of students' activity during the course to their final grade. The common opinion exists, that if a student is active during the lectures and is interested in the course news, his/her results should be higher. During the research it was stated that the data shares-out to the 3 clusters.

This clusterisation results are presented in the picture 3.4.
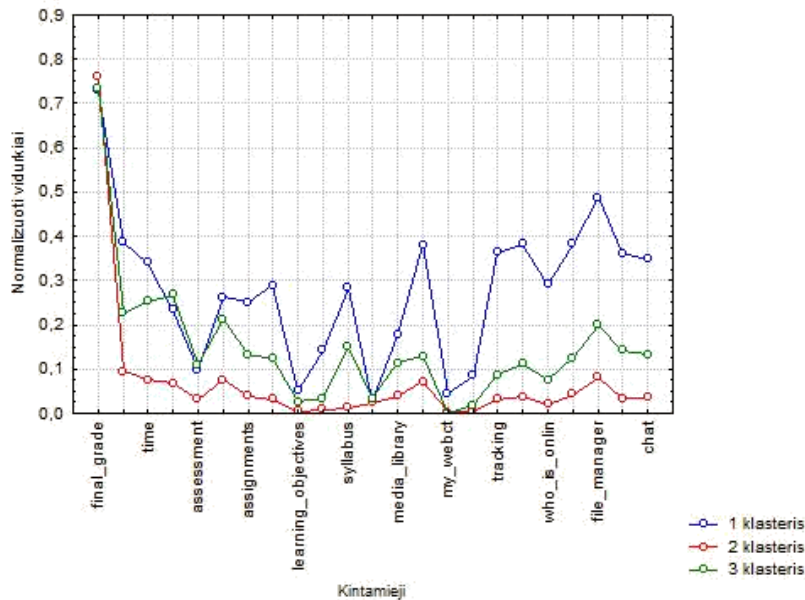
Picture 3.4. Students' activities clusterisation using k-means algorithm

For the deeper data analysis students' number in each cluster was calculated as well as the average of the final grade. Those results are presented in the table 3.3.

Cluster	Students number	Students number (%)	Final grade
1	44	8,33333	7,568182
2	360	68,18182	7,836111
3	124	23,48485	7,604839

As we can see the final grade of each cluster differs only slightly therefore willing to obtain more information each cluster was analysed separately.

3.5.2. Personalised „tasks' bag" formation using multidimensional association rules

The search of logical relations in the data is one of the common datamining tasks. Logical rules can explain very complicated relations. Logical rules allow to make prognosis and help to relate individual life areas into one unit (Sakalauskas 2008). There is a "shopping basket" task widely spread.

Let $I = \{i_1, i_2, \dots, i_m\}$ is a set of possible purchase, called elements. Association rules are expressed in form $A \subseteq B$, where A and B are all set elements complying conditions $A \subset I$, $B \subset I$ and $A \cap B = \emptyset$. The rules $A \Rightarrow B$ support is s if $s\%$ transactions from I are $A \cup B$, $\text{supp}(A \Rightarrow B) = \text{supp}(A \cup B)$. A frequency in which B follows from A is called rule's *confidence c*.

Multidimensional association rules

Associative rules having two or more dimensions inside are called *multidimensional association rules*. In picture 3.5. the case is presented when the food can have several dimensions. In this case the bread purchased can be either rye, or white.
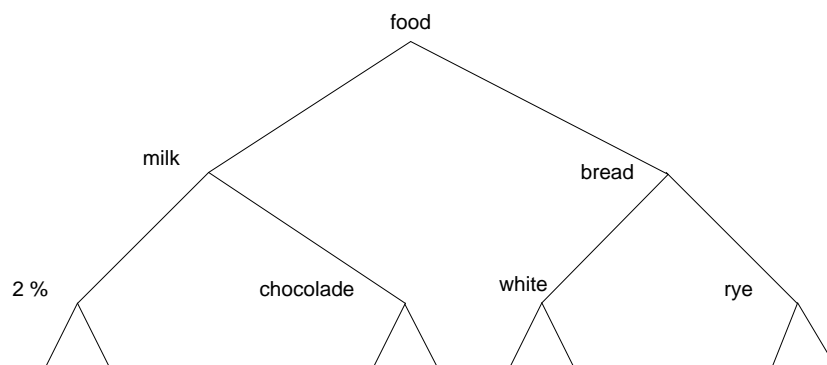
Picture 3.5. „ Shopping basket" in the structure of tree

Using similar principle it is possible to transcode the current students evaluations and prepare data for the algorithm ML_T2L1.

In the research three tasks of different topics were used for the students testing. Those topics were encoded with the appropriate variables (table 3.4).

Table 3.4. Decoding of the tasks topics

Topic	Variable
MS Excel. Financial functions	x1
MS Excel. Optimisation tasks	x2
SPSS. Linear regression	x3

For the association rules search categorical variables are used, therefore students' evaluation is transcoded. The result of this transcoding is indicated in the table 3.5.

Table 3.5. Transcoding of student's grades to 4 categories

Evaluation interval	Change
0–4	Negative
5–6	Weak
7–8	Good
9–10	Excellent

As one topic can have up to four different evaluation, *multidimensional association rules* have been searched in the research. For that ML_T2L1 algorithm was used which is dedicated for searching *multidimensional association rules*. To use this algorithm it was necessary to transcode obtained data once again (Picture 3.6.).
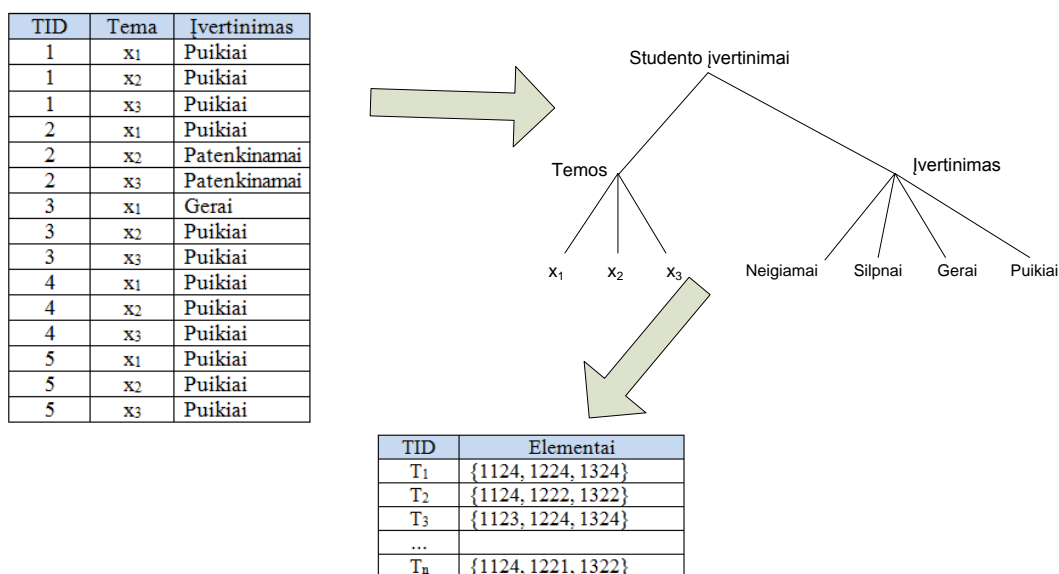
Picture 3.6. Data preparation for using ML_T2L1 algorithm

After the transcoding performed and putting all the data in one table it is possible to form students' tasks „baskets", and using ML_T2L1 algorithm the following rules can be found (table 3.6.).

Table 3.6. „Tasks basket" making rules

Rules	Support (≥20 %)	Confidence (≥50 %)
(x1, Good) ∧ (x2, Average) ⇒ (x3, Weak)	0,2	0,6
(x1, Average) ∧ (x3, Weak) ⇒ (x2, Negative)	0,2	0,55
(x2, Average) ∧ (x1, Average) ⇒(x3, Weak)	0,2	0,67

The obtained results indicate that having students' previous evaluations according to the separate topics we can form such "tasks baskets" with the big probability that students cannot skip any of the topics (or the aims stated).

Based on the data obtained, the scheme of such "tasks baskets" preparation, publication and integration into the VLEs was formulated (picture 3.7.).

The idea of this module is that collecting students' evaluations from VLE Moodle, transcoding them according to the separate topics, applying datamining methods and using computer agents, we can formulate personalised "tasks baskets" for every students.
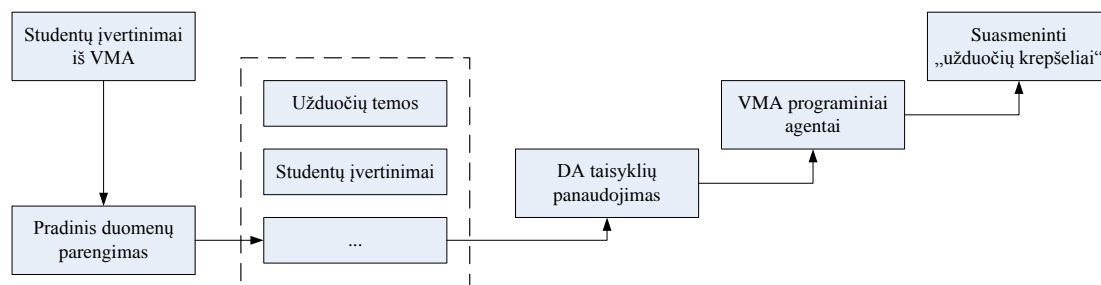
Picture 3.7. Usage of multidimensional association rules for the personalised „tasks baskets" preparation for VLE users (Preidys 2010a)

Presentation of the personalised tasks simplify observation of students activities for a teacher and ensure, that students would achieve the aims and objectives of the course.

4. Learnograms and their usage in the learners' activities evaluation

In the dissertation learnograms are introduced – the tool for learners' activities visualisation. Learnograms are still not used in VLEs, yet.

4.1. VLE visualisation means

In the distance studies the direct student – teacher communication is very limited. Such communication usually occurs only at the beginning of the studies and during the final evaluation. But willing to evaluate student's knowledge and skills objectively the teacher should know each student's activities during all his/her studies period. For this purpose students' activities tracking tools are used in VLEs. But those tools present only the statistical information which is not enough to fix the peculiarities of the study process.

4.2. Learnograms definition

As learnograms are a very new issue, they are rarely met in the literature. The author's of the dissertation suggested the following definition: **Learnograms** are generalised representation of learning process related to the evaluations.

Some authors define learnograms as follows: learnograms are visual learning variables representation during the time intervals. In such diagrams X axis contains time periods and Y axis – learning variables. Learning variables are student's logins to the particular learning object. Nevertheless, there are much more variables, such as the logins to the VLE, VLE objects, the order of access to the objects, time spent for the separate objects, etc. There are even more advanced objects, which are deduced from those accumulated VLE variables, for example, learning strategies, learning efficiency, etc.

In picture 4.1. a learnogram is presented where a tutor can find dynamics of student's logins during the whole study process. In total, a tutor can find the analysis of activities during a fixed period of students' group consisting of three students.
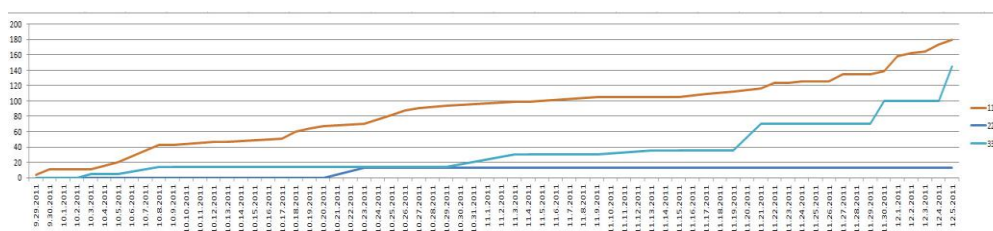
Picture 4.1. Dynamics of students' 111, 222 and 333 logins during the selected period

Group work is not a new idea, but it's still being introduced to the Lithuanian universities, colleges and even secondary schools. The most challenging issue is to

evaluate each student's input to the whole project. Moreover, this is especially relevant using VLEs and distance education.

Learnograms help to evaluate an input of every single group member to the common project. Of course, learnograms have a recommendation nature but still are very helpful for the teacher to prepare to for final evaluations. In picture 4.1. it is possible to notice that student *111* worked in a consistent way periodically logging the VLE and performing various activities. Thus, it is possible to assume that he was the key player in the group. Student *222* started to study a little bit later but also was very active and made his input to the common group project. However, student *333* practically didn't perform any activities and, as it happens, joined the group only before the final presentation.

4.3. The results of learners' activities evaluation

First of all the concrete student was distinguished and his activity in periods of the days was presented using graphical means (picture 4.2.). The Y axis has no data because the activity was expressed by the circle size. It is notable that student's activity was decreasing during the course and at the end of the course, he almost stopped to use VLE.
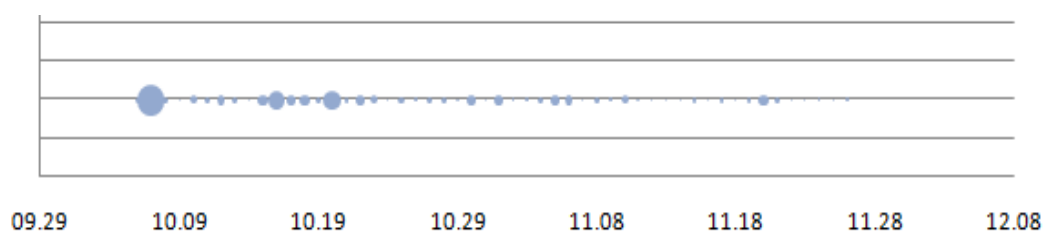

Picture 4.2. Student's 55161762001 activities in VLE in the period of days

Willing to identify the reasons for decreasing the student's activity, it is necessary to analyse the study process more carefully. One of the possible reasons is decrease in tutor's activities. From the other learnogram, which indicates student's activities by learning variables (picture 4.3.) it is notable that at the end of the course the student only sent his final assignment (day 11-30, variable *assignments*).
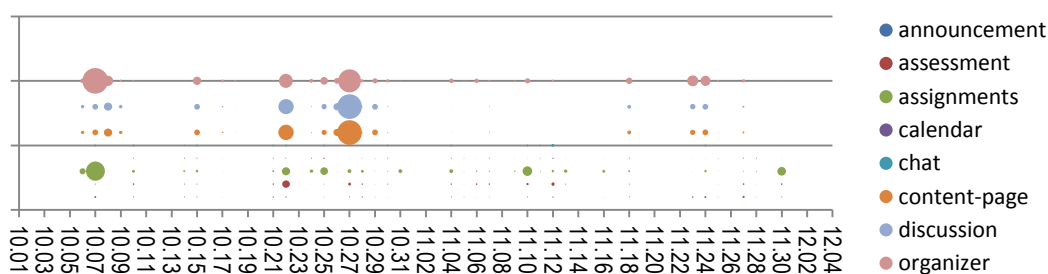

Picture 4.3. Student's 55161762001 activities in VLE according to the days and learning variables

From the data of the next learnogram (picture 4.4.) it is possible to conclude that with the exception of several days in general the student spent only short but frequent periods in VLE.
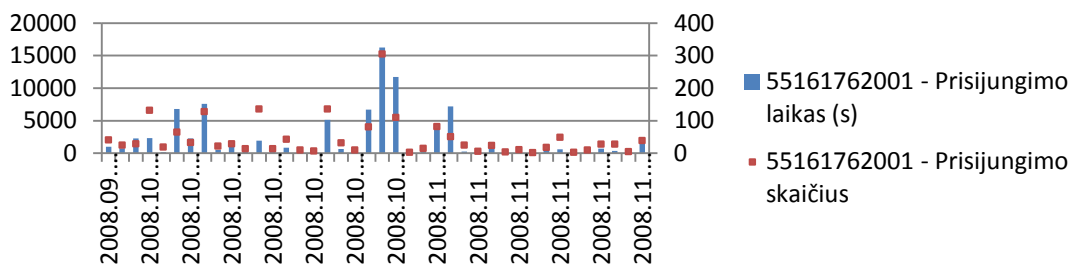
Picture 4.4. Number and duration of student's 55161762001 logins to VLE

While analysing the student's level of knowledge, a number of successful self-evaluation tests was counted in the research. In the concrete course every topic had self-evaluation tests and students could check their knowledge unlimited number of times. There were 10 self-evaluation tests in total in the course. In the picture 4.5. a specific student's self-evaluation tests performed are indicated and the number of trials is expressed by the radiuses of bubbles. It is notable that from 9th till 27th day the student didn't do any of self-evaluation tests, but on the 28th day he/she performed even 5 tests, once each. On the 9th day the test no. 4 was the most difficult one, and the positive evaluation was achieved only on the 5th time.
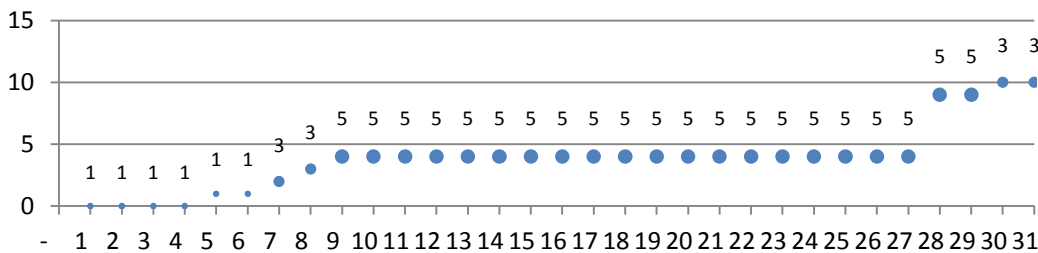
Picture 4.5. Self-evaluation tests performed and number of trials of the student 55161762001

Using the learnograms not only the students work during all their study period can be predicted, but also their learning styles can be analysed, students' activation steps can be foreseen, etc.

5. Personalisation of learners' styles

Many learners' styles are very different. Some of them prefer to listen and to speak while the others need to analyse the text or learn using visual aids. Nevertheless, the most of learners use mixed style. In this dissertation the author, using Honney and

Mumford's typology, analyses learners' styles identification and visualisation by examining students' activities in VLE and applying datamining methods to the data accumulated.

5.1. An overview of learners' styles

Before planning the prepare and submit a distance learning course authors should pay attention to the fact that students study in different ways: some start reading the given materials one after the other, the others only pay attention to the unclear cases, the third group would rather choose virtual discussions, etc. Therefore, after analysing learning activities and identifying learner's style it is possible to present a personalised learning materials and to choose better methods of course presentation.

In the traditional teaching process teachers or tutors can easily observe learner's efforts in order to determine if a learner is studying all the period and to establish the learners' style. Nevertheless in a distance education and using VLEs this becomes more difficult.

There are a lot of learning style models described in the literature. The main widely used and analysed learning styles are of Colb, Honney and Mumford, Pasc, Felder and Silverman.

5.2. Honney and Mumford learning style model

Currently in the learning styles research Honney and Mumford typology is predominating. Those authors have identified four learning styles each of which is related to the priority to the concrete learning cycle stage (picture 5.1.)
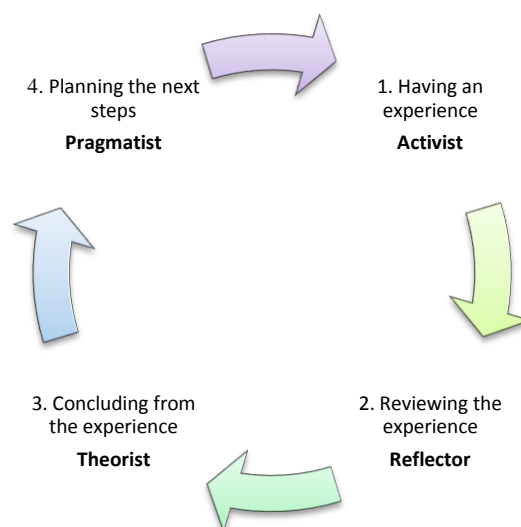
Picture 5.1. Honney and Mumford learning style model

5.3. Clusterisation and adaptation of the logical rules derivation for personalisation of the learners styles

For the learning styles identification the tested group had to pass learning style identification test prepared by Honney and Mumford's methodology. The test consisted of 40 questions which were also published in VLE VMA Moodle. 60 students of the third year of Business and Advertising management participated in the research. Those students already studied in VLE for almost a year therefore quite a lot of information about students' activities in VLE was accumulated.

As mentioned before, after performing the necessary calculations aggregated data is transferred to the MS Excel 2007 (picture 5.2.). In this table it can be observed how actively students participated in the study process, which learning resources they used in their studies and which learning style to they were assigned according to the results of testing method.

Student	assignment	blog	course	forum	glossary	questionnaire	quiz	resource	upload	user	Style
st_4286	0	0	2	0	0	9	1	0	0	0	R
st_4285	0	0	15	8	0	6	3	5	0	0	A
st_4280	0	0	8	2	0	6	18	0	0	0	R
st_4279	0	0	14	0	0	7	33	6	0	0	T
st_4278	0	0	10	2	0	7	39	17	0	0	R
st_4276	0	0	6	0	0	18	18	26	0	0	T
st_4275	0	0	9	0	0	7	18	4	0	0	P
st_4272	0	0	5	0	0	12	19	0	0	0	R
st_4270	0	0	11	0	0	7	24	22	0	0	R
st_4269	0	0	8	14	0	6	36	10	0	0	R
st_4268	0	0	10	0	0	8	15	5	0	0	R
st_4266	0	0	5	0	0	10	31	23	0	0	R
st_4264	0	0	4	0	0	6	20	2	0	0	A
st_4263	0	1	9	13	0	9	36	19	0	2	A
st_4262	0	0	22	8	0	17	16	6	0	0	R
st_4261	0	0	14	11	0	9	35	3	0	2	P
st_4260	0	0	7	0	0	10	21	3	0	0	A

Picture 5.2. Data on activities in VLE of students whose learning styles were identified using Honney and Mumford questionnaire

In the next stage, using the same algorithms for the data accumulation and preparation for datamining methods, the data of activities in VLE of students who didn't participate in the learning styles identification test was accumulated. This data is moved and kept in the other MS Excel table. The difference between those tables is only that the first table indicates every student's learning style, while newly formed table has no information about the learning styles (picture 5.3.).

| Student	assignment	blog	course	forum	glossary	questionnaire	quiz	resource	upload	user	Style
st_386	0	0	16	0	0	0	47	5	0	1	
st_3858	0	0	15	0	0	0	515	39	0	3	
st_385	24	0	36	4	0	0	104	32	0	5	
st_3839	0	0	36	0	4	0	63	40	0	1	
st_3835	0	0	30	0	1	0	667	153	0	0	
st_382	15	1	148	5	0	0	161	48	0	10	
st_381	0	0	26	3	0	0	125	5	0	2	
st_3805	0	0	16	0	0	0	397	82	0	2	
st_380	2	1	65	3	0	0	83	19	0	18	
st_379	0	0	11	0	0	0	36	20	0	4	
st_3755	0	1	4	0	6	0	1	30	0	2	
st_3747	0	0	5	0	0	0	10	38	0	0	
st_3741	0	0	78	0	2	0	2353	104	0	26	

Picture 5.3. Data about activities in the VLE of students which participated in the automatic learning style identification

Before starting to apply datamining tools for this data, two tables were merged into one. The calculations were performed using MS Excel tool MS SQL 2008 Datamining Add-Ins for MS Excel 2007.

5.4. Case-based-reasoning systems

Performance principles of *Case-Based-Reasoning* systems used in this research seam very simple from the first glance. Those systems find a close analogue to the data sample and select the response (precedent) which was true for the analogue.

The CBR cycle is used while solving tasks (picture 5.4.)
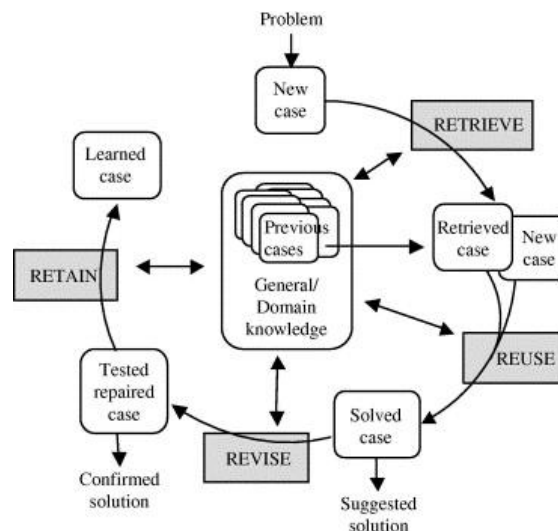
Picture 5.4. CBR cycle

After the analogue has been found and compared to the analogues in the database, the next task is to find an optimal solution. Very often, the nearest neighbour method is used for searching the optimal solution. The essence of this method is that the most accurate precedent is looked for in the test data of the database; and it can be either numerical, text, or logical value). In such calculation similarity function (metrics) is

derived using all the correlation features. The distance metrics are usually used to measure the distance between the features:

Euclidus metrics (for the quantitative variables):

$$d_{ik} = \left(\sum_{j=1}^{N}(x_{ij} - x_{kj})^2\right)^{\frac{1}{2}}, \qquad (3)$$

Heming metrics (for the nominal/qualitative variables):

$$\mu_{ij}^H = \frac{n_{ik}}{N}, \qquad (4)$$

here n_{ik} is a number of coinsidating features in the sets of x_j ir x_k.

Mahalanoby metrics (for quantitative variables):

$$d_{ik}^M = (x_{ij} - x_{kj})^T W^{-1}(x_{ij} - x_{kj}), \qquad (5)$$

here W – covariation matrix $X = (X_1, X_{21}, ... X_n)$.

Nevertheless in practice while calculating similarities between the precedents, the weighting factor can be assigned to the features, because different features can influence decision making in various cases. Therefore, in general precedent distance is measured by the following formula:

$$\sum_j w_j \times sim(x_{ij}, x_{kj}), \qquad (6)$$

here w_{ij} j feature weight, sim – similarity function, x_{ij}, x_{kj} the value of current precedent and feature. The weights were not used during the research – it was assumed that they are 1 for the all precedents.

5.5. Results

After a set of examined students' activities in VLE was analysed with datamining tool, the identified models were applied to the main students set and every student was assigned to the particular learning style, filling in the column *Style*. The models found out are presented in the picture 5.5.

Column	Value	Favor	Relative Impact
resource	23,562 - 62,532	A	
course	19,978 - 44,230	A	
forum	0,000 - 0,000	A	
user	0,000 - 0,000	A	
forum	0,000 - 4,037	A	
assignment	2,931 - 10,865	P	
course	2,000 - 5,910	P	
forum	0,000 - 0,000	P	
forum	0,000 - 4,037	P	
questionnaire	6,000 - 6,284	P	
quiz	0,000 - 9,696	P	
resource	0,000 - 0,956	P	
user	4,977 - 15,113	P	
user	2,037 - 4,977	P	
forum	8,645 - 24,531	R	
questionnaire	13,198 - 25,116	R	
quiz	46,637 - 110,318	R	
assignment	2,931 - 10,865	T	
assignment	0,630 - 2,931	T	
resource	12,259 - 23,562	T	
questionnaire	6,000 - 6,284	T	
questionnaire	6,284 - 9,741	T	
quiz	0,000 - 9,696	T	
quiz	9,696 - 28,167	T	
resource	23,562 - 62,532	T	

Picture 5.5. Visual results received after using *Fill From Example* data analysis method

While analysing the models received it is possible to notice that students identified as *activists*, login to their VLE course most times to compare to other students. The logins number is in the interval of 19,978 – 44,230. That proves their active learners style. On the other hand, not only the login number indicate their activity – those students were most interested in learning resources as well.

In the activities of *pragmatics* the stability can be observed. Their logins are not so frequent, but they try to find practical ways of VLE objects' usage, perform *assignments*. But they usually escape from reading learning materials, as try to achieve all the results in a practical manner.

Reflectors are not active. They use independently accumulated data and the data received from the other people for their thorough reflection. That is obvious from the models also – reflectors' activity is based only on the usage of communication objects. Reflectors observe the discussions, the results of questionnaires; also very often they observe their tests results and analyse their own mistakes.

Theoretic refer to the observations and reflection of the experience. From the results obtained it is possible to notice that they are interested also in learning resources, observe and analyse assignments and tests.
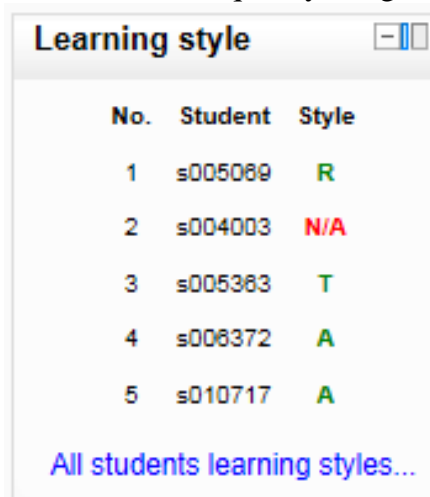
5.6. Practical application of the research results

The results of this research were applied practically. From the 20th of March till the 2nd of April 2011 an Erasmus Intensive programme project WISDOM'2011 – *Web Information System Data Organisation Modelling* was implemented. 60 students from 10 European countries have participated. For two weeks the students worked in groups analysing WEB data and applying datamining methods to them. Before coming to Lithuania students had to prepare some homework, which assignment was published in VLE Moodle. Project coordinators selected all students' activities data before students' coming to Lithuania and, using learning styles models, identified the students' learning styles. Face-to-face working groups were formed according to the results obtained. During the project activities all groups were observed by the tutors, and again the students' activities were recorded. The result of this practical application turned very prospectively – around 70 percent of the activists became leaders of their groups: started to distribute the assignments, offered methods for information search, were extremely active in the social-cultural activities.

The representatives of the other learning styles matched their styles in 30 – 40 percent. The less matching can be explained this way: wiling to establish other styles it is necessary to analyse the login time to a concrete VLE resource together with the activities during this login. This will be done during further investigations.

After the end of this project when the questionnaire about the quality of group composition was given to all the participants, only a small percentage (5 – 8 %) of respondents mentioned, that working groups were formed in a wrong way. Some of them indicated other reasons for disliking the group formation, for example, that they needed to search for info in the websites of non-English speaking nationalities.

One more practical learning style application is represented in the picture 5.6. Here the programme agent is represented which is integrated into VLE Moodle. This agent, after analysing users' activities

Picture 5.6. Recommended module for learning style identification, integrated into VLE Moodle

during their studies, identifies students' learning styles and offers the teacher the main issues, which should be taken into account while presenting materials to a concrete student[1].

6. Creation of software prototype and integration of separate modules to the open source VLE Moodle

In this chapter two systems will be presented, which, after analysing students' activity data, not only present statistical diagrams, but also, using the datamining methods, give the further information: forecast the number of students' logins, analyze tests, indicate the most difficult and easy questions for the tutor, thus recommending to transform test, cluster students according to their activities in VLE, etc.

The first system was realised during the dissertation writing period. It was designed and created separately from VLE Moodle. This system took data necessary for the analysis from VLE Moodle database, prepared it for the datamining methods application and presented the results to the users.

The second system will be created while implementing national project of Lithuanian Distance Education Consortium „LieDM network development". During this project the other prototype implementation will be chosen – the integration of separate modules to VLE Moodle.

6.1. Prototype creation

Students activities were selected using eight sub programmes created using PHP programming language and transferred to the other MySQL database. Each of those sub programmes selected different data: information about test results, users' logins and their activities in VLE data, users time spent in each course module, etc. Those sub programmes perform calculations needed and transform current data to the format applicable for datamining methods. Prepared data goes to the other MySQL database where it is kept for the further data analysis.
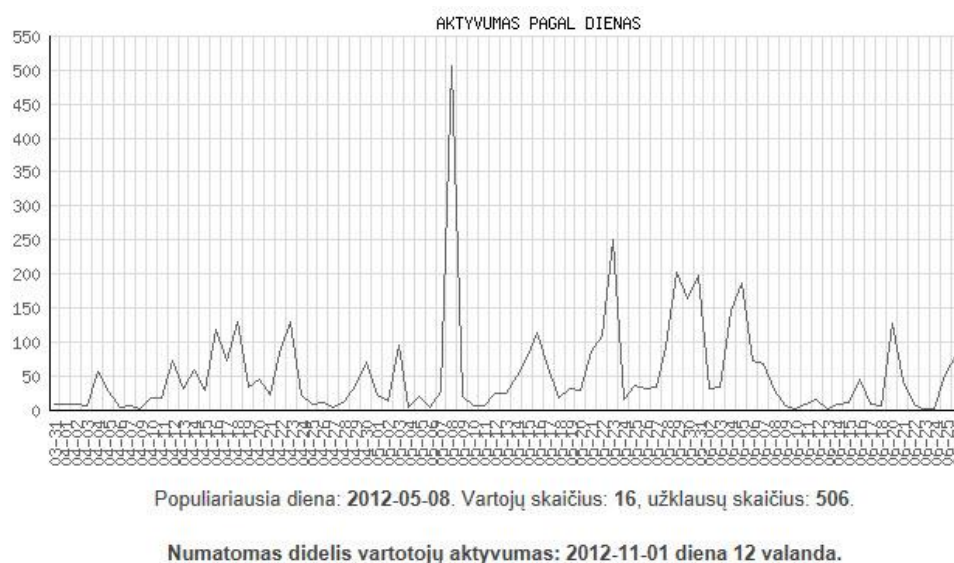
---

[1] N/A means that there is not enough data yet for the student's learning style identification.

## 6.2 Visualization, analysis and interpretation of the results obtained

After programming all the sub programmes all of them were merged to one system. In this system users were categorised to the same categories as in VLE Moodle. They can login to this system with the same login data as to the VLE Moodle.

The presented information depends on the user's category. System administrator can see information actual for him/her, like a course size, place taken in the server, server load flow, number of accesses. He/she can see the forecasted users' activity also (picture 6.1.). Taking this information into account the server administrator can plan server load flows, inform teachers about possible server disorders, etc.

Populiariausia diena: 2012-05-08. Vartojų skaičius: 16, užklausų skaičius: 506.

Numatomas didelis vartotojų aktyvumas: 2012-11-01 diena 12 valanda.

Picture 6.1. Information about server load and forecasted users' activity

The most information is dedicated for the course tutor and course author. The system informs teacher about the inactive students, unused or very rarely used tools or resources, informs about too easy or too difficult tests questions, allows to track each student's performance (picture 6.2.) and activities in the module (picture 6.3.).
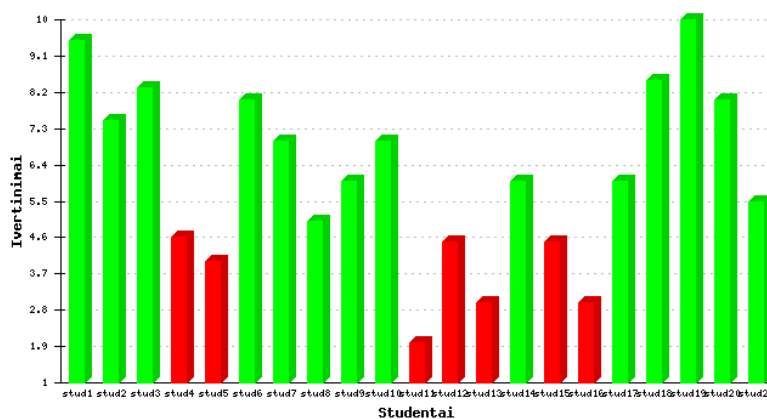

Picture 6.2. An average of students' evaluations
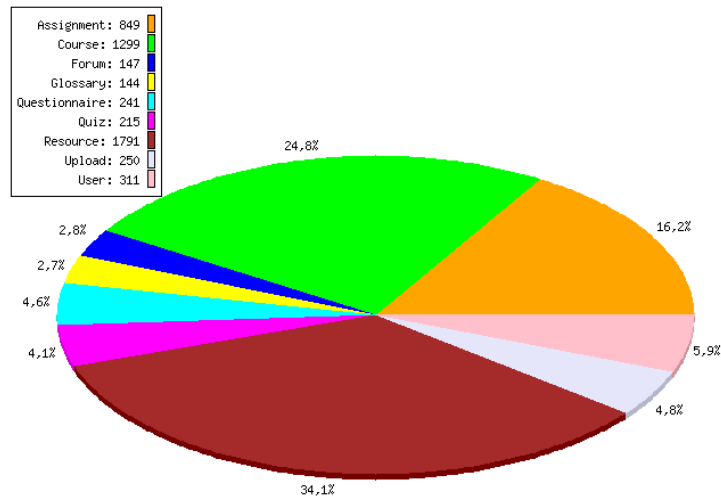
Picture 6.3. Student's time spent the module according to the separate objects

After analysing users' activities in VLE and applying clustering method, the users are divided into 3 clusters. In the first cluster the active users are those, who frequently use a lot of Moodle tools, actively participate in the discussions, are interested in teacher's new assignments and course news. The third group are passive users. Those are the users which rarely login to VLE, spend there only some time, are not interested in active performance and usage of Moodle tools (picture 6.4.). The second group are users who are not very passive but, on the other hand, are not spending a lot of time for the analysis of Moodle learning objects.
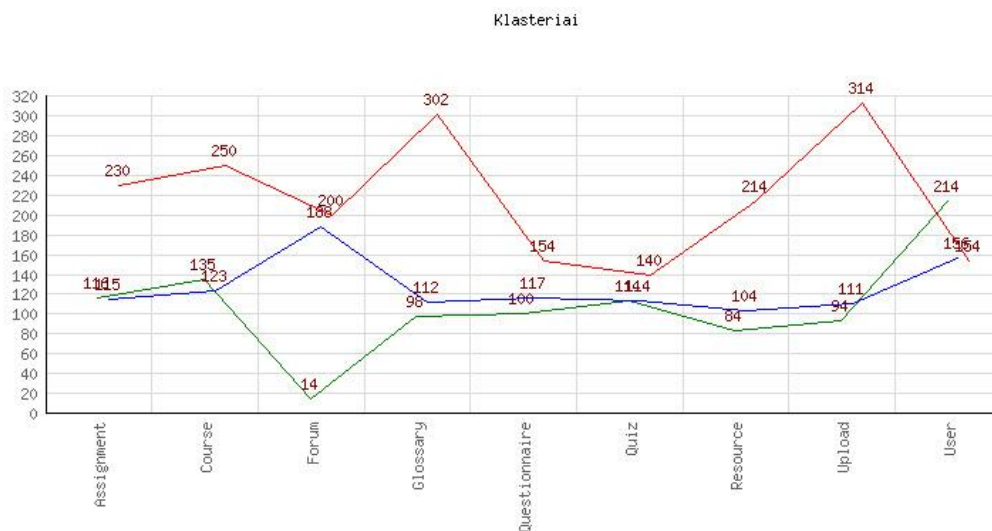
Picture 6.4. Users clusterisation according to their activities in VLE

The other interesting information is students' distribution according to their activities in VLE and their final evaluation (picture 6.5.). This allows us to observe how

module resources impact the final grade, i.e. it is possible to analyse which activities are taken by the students with the best results and which activities by those with the negative achievements.
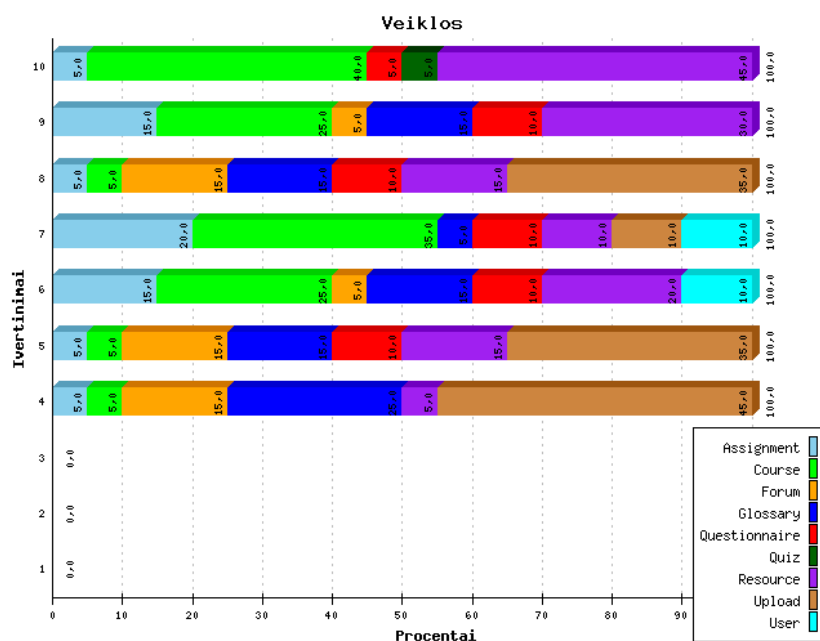
Picture 6.5. Students' activities according to the final evaluation and activities in VLE

6.3. Integration of software modules into VLE Moodle

The aim of LieDM project „LieDM network development" is an improvement of LieDM network functionality and quality of service purposefully using the financial resources. Using project results it is aimed to increase the possibilities of all institutions participating in LieDM network offer contemporary requirements conforming studies in an electronic way, to expand the volume of services and their accessibility, to increase users' network. The project will be implemented from January 2013 till April 2014.

There will be 4 main activities in this project:

1. Expanding possibility of video conferences.
2. Implementation of quality assurance.
3. Expansion of VLE functional possibilities.
4. Dissemination of project results.

While performing the third activity the research of this dissertation author was already used and several computerised modules have been designed for Moodle, which will be developed into programmes at the later stages. With the help of these modules standard Moodle equipment will be increased with the following modules:

36

1. The module of users' activities analysis which analyses users' activities, send actual information to the course tutors and students, reply to the letters, activate passive course participants.
2. A module of e-learning quality evaluation which evaluates the quality of course preparation, give suggestions for the course quality improvement, summarised reports about the quality of VLE courses delivery.
3. Visualisation module which present statistical diagrams and learnograms of fixed volume and details.
4. A module of users' learning style establishment and learning process recommendation module.

Conclusions

The accomplished research revealed personalised learning possibilities in the distance education using Virtual Learning Environments. Based on the results of the research it is possible to make the following conclusions:

1. In the electronic learning systems direct contacts between the teacher and the learner are very limited, therefore it is actual to create architectures of electronic learning warehouses adapted to personalised learning and using datamining methods.
2. Five types of distance education users are identified in the research: teachers, tutors, students, administrators and auditors, and their needs and activities while carrying out distance education are described in details.
3. The clusterisation method used in the dissertation allowed to identify 3 learners' clusters, and deeper investigation established an impact of activities of those clusters representatives to the final learners' knowledge evaluation. The research performed allows to state that personalised learning materials should be presented for representatives of the different clusters.
4. Using Honney and Mumford's learning style identification typology, the method of identifying student's learning style in an automatic way was created using *Case-Based Reasoning* algorithm. During the experiment the right learning style was identified for even 70% participants.
5. Description of the new summarised learning process representation, related to the final evaluation, was presented. It was established that the usage of learnograms in distance education allows to receive reliable feedback from the

distance education learners. While analysing student's learnogram, a teacher can see student's workload during the whole study process. This enables to reduce the possibilities of students' plagiarisms and other unfair activities.

6. A software prototype for analysis of learners' activities was created, which performs separately from VLE Moodle. An advantage of this prototype is possibility to use it for another than Moodle VLEs' users' activities analysis.

7. The students' learning style identification tool is prepared and integrated into VLE Moodle, which allows to inform teacher about every single student's learning style and give recommendations, based on peculiarities of this style.

8. The results of this dissertation research will be used implementing LieDM project „LieDM network development" and Moodle system will be supplemented with computerised agents for teachers and system administrators.

Publications

1. Žilinskienė I., Preidys S.; 2012. A Model for Personalized Selection of a Learning Scenario Depending on Learning Styles, // (Accepted Databases and Information Systems VII, Frontiers in Artificial Intelligence and Applications Series of IOS Press, 2012-08-30)

2. Preidys S., Žilinskienė I.; 2012. Nuotolinio mokymosi kurso personalizavimo modelis mokymosi veiklų atžvilgiu // Elektroninis mokymasis, informacija ir komunikacija: teorija ir praktika, Vilniaus universiteto Elektroninių studijų ir egzaminavimo centras, psl. 111-132, ISBN 978-609-459-030-6.

3. Preidys S., Sakalauskas L., 2012. Išmaniųjų modulių integravimo į VMA Moodle galimybės: nuo teorijos prie praktikos. //Mokslo taikomųjų tyrimų įtaka šiuolaikinių studijų kokybei V: ISSN 2029-2279

4. Preidys S., Sakalauskas L.; 2011. Nuotolinio mokymosi stilių personalizavimas, // Informacijos mokslai 56: psl. 42-49, ISSN 1392-0561.

5. Preidys S., Sakalauskas L. Analysis of Students' Study Activities in Virtual Learning Environments Using Data Mining Methods // Technological and Economic Development of Economy. Vilnius: Technika, 2010, Vol. 16, No. 1, p. 94-108. DOI: 10.3846/tede.2010.06, ISSN 1392-8619

6. Sakalauskas, L.; Preidys S.; 2009. Nuotolinių studijų vartotojų poreikių analizė, // Informacijos mokslai 50: psl. 117-123, ISSN 1392-0561.

Briefly about the author

Saulius Preidys graduated from Paįstrys high school in Panevėžys district in 1989. In 1995 graduated from Vilnius Pedagogical University and was awarded the Master Degree in Informatics. Doctoral student at the Institute of Mathematics and Informatics from 2008 to 2012. Works as a director of E-Learning and Examination Centre in Vilnius University. A member of Lithuanian Computer Society, National Association of Distance Education, Lithuanian Distance Education Consortium Board. E-mail: saulius.preidys@esec.vu.lt.

Reziumė

Tyrimų sritis ir problemos aktualumas

Šiuolaikinės informacinės technologijos taikomos įvairiose srityse: versle, žiniasklaidoje, kasdieniame gyvenime, mokslo ir švietimo sektoriuje. Jau mokyklose mokiniai naudoja asmeninius ir planšetinius kompiuterius rinkdami informaciją, rengdami pristatymus, bendraudami tarpusavyje. Dar neseniai mokyklose, universitetuose buvo plačiai taikomi antros kartos žiniatinklio (WEB 2.0) metodai, tokie kaip WIKI, BLOG, socialiniai tinklai, tačiau jau šiandien prasideda trečios kartos duomenų (WEB 3.0) era: sistemos tampa „išmanios", interneto vartotojai gauna suasmenintą, jo poreikiams pritaikytą informaciją. Sistemos stebi ir analizuoja vartotojo veiksmus ir, reikalui esant, pasiūlo reikiamus kelius. Tai labai plačiai naudojama išmaniose paieškos sistemose (Yahoo!, Google, Hakia), semantiniuose socialiniuose tinkluose (Facebook, GroupMe!, Twins), elektroninės prekybos sistemose.

Mokymasis visą gyvenimą tampa būtinybe spartėjant gyvenimo tempui ir didėjant informacijos srautams. Pasaulyje populiarėja nuotolinis mokymas pasinaudojant programine įranga ir virtualiosiomis mokymo aplinkomis (VMA), leidžiančiomis studijuoti nuotoliniu būdu. Pasitelkus VMA galima mokytis vartotojui patogiu laiku ir tempu, vartotojas tampa nepriklausomas nuo studijų ir gyvenamosios vietos.

Plačiai naudojamos virtualaus mokymo aplinkos, tokios kaip BlackBoard Vista, Moodle, Doceos ir kitos suteikia galimybę besimokantiesiems ne tik studijuoti pateiktą medžiagą, tačiau ir atlikti savikontrolės testus, dalyvauti diskusijose ir pokalbių kambariuose, pildyti tinklaraščius (BLOG), skelbti savo mintis WIKI aplinkoje. Kita naudinga VMA funkcija – vartotojų veiksmų stebėsena. Kiekvienas vartotojo veiksmas

yra fiksuojamas VMA duomenų bazėse ir (arba) žurnalų bylose. Kurso dėstytojai ir kuratoriai gali matyti vartotojų veiksmų statistiką, analizuoti jų veiksmus, reikalui esant, priimti atitinkamus studijų eigos pakeitimus.

Stebėjimo įrankis yra integruotas praktiškai į kiekvieną VMA. Tačiau šie įrankiai atskleidžia daug mažiau informacijos, negu yra sukaupta. Virtualaus mokymo aplinkų administratoriai dažniausiai pašalindami šiuos duomenis iš serverių kaupiklių praranda daug naudingos informacijos. Analizuoti duomenis galima įvairiais statistikos metodais, tačiau kai duomenų kiekis tampa didelis, tenka taikyti duomenų tyrybos (angl. *Data Mining*) metodus, t.y. klasterizavimą, klasifikavimą, vizualizavimą, neuroninius tinklus ir pan.

Šiame darbe yra tiriami virtualaus mokymosi aplinkose sukauptų besimokančiųjų veiksmų duomenų parengimo taikyti duomenų tyrybai uždaviniai, duomenų tyrybos metodų integravimas į atviro kodo virtualaus mokymo aplinką Moodle ir šių duomenų vizualizavimas.

Naudojant nuotolines studijas, studentų mokymas, jų priežiūra ir vertinimas labai skiriasi nuo tradicinio mokymo. Kuratorius daug laiko praleidžia konsultuodamas studentus, skelbdamas metodinę medžiagą virtualaus mokymo aplinkose, taisydamas atliktus studentų darbus, dalyvaudamas bendrose asinchroninėse bei sinchroninėse diskusijose. Tačiau visa ši veikla atliekama studentui nutolus nuo savo kuratoriaus. Studentui atvykus galutiniam atsiskaitymui, kurso kuratoriui sunku nustatyti jo įdirbį per visą kursą, nustatyti jo mokymosi klaidas, pastebėti plagijavimo atvejus. Norint kuratoriui atsakyti į tokius klausimus tenka studijuoti studentų veiklą per visas studijas, atlikti įvairių duomenų analizę, priimti sprendimus. Studentų veiklos registravimo moduliai – tai viena svarbiausių VMA sudedamųjų dalių. Tačiau jų vizualizavimas yra nepakankamai išplėtotas praktiniam panaudojimui. Yra daug papildomų programų, skirtų vizualizuoti sukauptiems duomenims, tačiau jos nėra integruotos į VMA. Šiame tyrime nagrinėjama nauja vizualizavimo sritis – mokymosi diagramos (angl. *learnograms*).

Planuodami rengti ir teikti nuotolinio mokymosi kursą, kurso rengėjai turi atsižvelgti į tai, kad žmonės studijuoja skirtingais metodais: vieni pradeda skaityti pateiktą medžiagą iš eilės, kiti peržiūri tik nesuprantamas vietas, treti persikelia į virtualias diskusijas ir pan. Todėl išanalizavus mokymosi veiksmus ir nustačius

besimokančiojo stilių, vėliau jam galima pateikti suasmenintą mokymosi medžiagą, parinkti geresnius kurso pateikimo metodus (Preidys, Sakalauskas 2010b). Toks mokymo organizavimas pagerintų studijų kokybę ir leistų pasiekti geresnių rezultatų. Mokydamasis tradiciniu būdu dėstytojas ar kurso kuratorius gali nesunkiai įžvelgti besimokančiojo pastangas, stebėti, ar studentas mokėsi visą laikotarpį, nustatyti besimokančiojo stilių. Tačiau nuotoliniam mokymui naudojant virtualaus mokymo aplinkas (VMA), tai padaryti yra sunku.

Mokymosi stilių nustatymas kurso pradžioje yra svarbus edukologinis elementas siekiant gerų studijų rezultatų. Mokymo stiliams nustatyti naudojami du metodai: bendravimo ir automatinis (Duff, A. 2004), (Dimitrios, Botsios 2007). Pagal pirmąjį metodą besimokantieji turi atsakyti į mokymosi stiliaus nustatymo anketos klausimus. Dėstytojas apdorojęs rezultatus priskiria besimokantįjį prie vieno ar kito mokymosi stiliaus. Nustatant besimokančiųjų mokymosi stilių antruoju metodu, atsisakoma savęs vertinimo apklausos anketos. Tai leidžia objektyviau vertinti tiriamuosius, nes išvengiama galimai neteisingų atsakymų. Naudojantis automatiniu mokymosi stilių nustatymo metodu yra analizuojami besimokančiųjų veiksmai, kaupiama informacija apie studentų elgseną, ieškoma panašumų tarp besimokančiųjų. Automatiniam besimokančiųjų mokymosi stiliaus nustatymui labai tinka virtualiose mokymo aplinkose (VMA) sukaupti vartotojų veiksmų duomenys.

Šioje disertacijoje sprendžiamos šios pagrindinės problemos:

1. Vartotojų veiksmų VMA parengimas taikyti duomenų tyrybos metodus.
2. Vartotojų veiksmų vizualizavimas ir vaizdus pateikimas virtualios mokymo aplinkos vartotojams: administratoriams, dėstytojams, kuratoriams ir studentams.
3. Vartotojų mokymosi stilių modelio nustatymas analizuojant sukauptus vartotojų veiksmų duomenis VMA.

Darbo tikslas ir uždaviniai

Pagrindinis tyrimo tikslas – sukurti metodiką duomenų tyrybos metodams taikyti suasmeninto mokymo nuotolinėse studijose.

Siekiant šio tikslo buvo iškelti ir sprendžiami šie uždaviniai:

1. Nuotolinių studijų duomenų saugyklų tyrimas parengiant duomenų tyrybos taikymo galimybes.
2. Besimokančiųjų veiksmų vertinimas naudojant mokymosi diagramas.

41

3. Besimokančiųjų mokymosi stilių išskyrimas.
4. Suasmeninto mokymosi metodų sukūrimas ir taikymas.
5. Programinio prototipo sukūrimas ir atskirų modulių integravimas į atviro kodo VMA Moodle.

Tyrimo objektas ir metodai

Disertacijos tyrimo objektas – duomenų tyrybos metodų pritaikymas virtualaus mokymo aplinkose sukauptiems besimokančiųjų veiklos duomenims apdoroti. Šis objektas yra aktualus intelektualiai duomenų analizei ar duomenų tyrybai, nes interpretavus gautus rezultatus galima parengti modelius ir remiantis jais pagerinti nuotolinių studijų kokybę.

Pagrindiniai tyrimo metodai taikomi disertacijoje – stebėsena, dokumentų analizė, eksperimentas. Tyrimo metu buvo dirbama su realiais Vilniaus kolegijos ir Vilniaus universiteto virtualaus mokymo aplinkose sukauptais duomenimis, stebimi studentai per paskaitas ir jų veiksmai naudojant nuotolinio mokymosi metodus.

Darbo naujumas ir jo aktualumas

1. Sistemiškai išnagrinėtas duomenų tyrybos metodų taikymas virtualaus mokymo aplinkose sukauptiems vartotojų veiksmų duomenims apdoroti.
2. Sudarytas virtualaus mokymo aplinkų vartotojų veiksmų klasterizavimo metodas pritaikytas mokymosi veiklų analizei.
3. VMA vartotojų veiksmų vizualizavimui panaudotas iki šiol mažai taikytas metodas – mokymosi diagramos (angl. *learnograms*), leidžiančios stebėti visą besimokančiojo veiklą ir pasiekimus per pasirinktą laikotarpį.

Darbo rezultatų praktinė reikšmė

Disertacijos rengimo metu atliktų tyrimų rezultatai bus panaudoti integruojant kompiuterinius agentus į virtualaus mokymo aplinką Moodle: pagal LieDM konsorciumo vieną iš projekto „LieDM tinklo plėtra" veiklų, suprogramuoti moduliai bus integruoti į VMA Moodle ir jais naudosis visi LieDM konsorciumo nariai, 14 Lietuvos universitetų, kolegijų ir kitų mokymo įstaigų (Preidys, Sakalauskas 2012).

Taikant automatinį apklausų metodą, remiantis analogijų paieška (angl. *Case-Based Reasoning*) bei Honney ir Mumford (Honney, Mumford 1992) mokymosi stilių klasifikacija, parengtas bei integruotas į Moodle VMA modulis, leidžiantis nustatyti besimokančiojo mokymo stilių. Remiantis šiuo metodu parengtas programinis agentas, teikiantis rekomendacijas kurso kuratoriui.

Darbo rezultatų aprobavimas

Disertacijos tyrimų rezultatai publikuoti 6 moksliniuose straipsniuose. 1 straipsnis patenka į užsienio mokslo leidinį, įtrauktą į Mokslinės informacijos instituto pagrindinių žurnalų sąrašą, 5 – į recenzuojamus Lietuvos ir užsienio leidinius.

Autorius dalyvavo ir pristatė rezultatus 11-oje respublikinių ir tarptautinių mokslinių konferencijų: Baltic DB&IS 2012 (Tenth International Baltic Conference on Databases and Information Systems), V mokslinė-praktinė konferencija „Mokslo taikomųjų tyrimų įtaka šiuolaikinių studijų kokybei –2012", KODI-2011 (Lietuvos kompiuterininkų sąjungos konferencija), tarptautinė konferencija „E. švietimas: mokslas, studijos ir verslas – 2010", LOTD-2010 (3-oji Lietuvos jaunųjų mokslininkų konferencija Operacijų tyrimai verslui ir socialiniams procesams LOTD – 2010), tarptautinė konferencija „*Innovation and creativity in e-learning*", KODI-2009 (Lietuvos kompiuterininkų sąjungos konferencija) ir kt.

Straipsnių ir konferencijose Vilniaus universiteto Matematikos ir informatikos instituto institucijos vardu skaitytų pranešimų sąrašas, pateikiamas darbo rezultatų aprobavimo skyriuje.

Disertacijos struktūra

Disertacinis darbas susideda iš įžangos, tyrimų srities aprašymo, šešių skyrių, skirtų rezultatų pristatymui, išvadų, cituotos literatūros sąrašo.

Pirmame disertacijos skyriuje pristatoma duomenų tyrybos metodų panaudojimo nuotolinėse studijose problema ir jos aktualumas, suformuluojamas darbo tikslas ir uždaviniai. Antrame skyriuje pristatomos šiuolaikinės nuotolinio mokymo architektūros, tokiose studijose naudojama programinė ir techninė įranga ir šiame procese dalyvaujantys veikėjai. Trečiame skyriuje apžvelgiami duomenų tyrybos metodai, jų pritaikymas įvairiose srityse. Didesnis dėmesys skiriamas duomenų tyrybos metodų pritaikymui nuotolinėse studijose, apžvelgiami taikymo etapai ir galimybės. Ketvirtame skyriuje pristatomas virtualaus mokymo aplinkose dar nenaudojamas duomenų vizualizavimo tipas, skirtas besimokančiųjų veiksmams analizuoti – mokymosi diagramos (angl. *learnograms*), praktinis jo pritaikymas. Penktajame skyriuje aprašoma besimokančiųjų mokymosi stilių analizė, duomenų tyrybos metodų taikymas automatiniam mokymosi stiliaus nustatymui. Šeštame skyriuje pristatomas realiai

veikiantis studentų veiklų analizavimo prototipas ir praktiškai aprašytas duomenų analizės modulis, kuris bus integruotas į VMA Moodle.

Glaustai apie autorių

Saulius Preidys 1989 metais baigė Panevėžio rajono Pajstrio vidurinę mokyklą. 1995 metais baigė studijas Vilniaus pedagoginiame universitete ir jam suteiktas informatikos mokslų magistro laipsnis. Nuo 2008 iki 2012 metų Matematikos ir informatikos instituto doktorantas. Dirba Vilniaus universiteto Elektroninių studijų ir egzaminavimo centro direktoriumi. Yra Lietuvos kompiuterininkų sąjungos, Nacionalinės distancinio mokymo asociacijos narys, Lietuvos distancinio mokymo konsorciumo tarybos narys.

El. paštas: saulius.preidys@esec.vu.lt.