Online learning behavior and web usage mining

PETER TOTH
Trefort Agoston Centre for Engineering Education
Obuda University
Nepszinhaz u. 8, Budapest
HUNGARY
toth.peter@tmpk.uni-obuda.hu

Abstract: The application of a virtual learning environment has become widespread in Hungarian higher education. Questions of quality and adaptivity are increasingly gaining dominance, manifested in course development and course management which takes the individual specialities of learners well into consideration. In order to increase the adaptivity of the learning process, we need to have exact and relevant information on the learner's learning characteristics and preferred learning strategies in an online environment. In this we may be aided by web mining methods, which process data from interaction between the learner and the learning objects. By the application of these methods we were trying to find an answer to the question whether any conclusions from the patterns of online learning activities were to be drawn as to preferred learning characteristics, methods and strategies and also whether the major variables of online learning behavior were possible to define.

Keywords: adaptive online learning environment, web usage mining, online learning behavior

1 Introduction

Therefore by the application of web mining methods as inductive examination procedures certain cognitive processes, strategies, learning characteristics are to be deduced and special learning habits and difficulties are to be isolated and typified. In the context of syllabus developer – tutor – electronic syllabus – learner two kinds of syllabus developing processes are collated. The learning process conceived and formed by the developer and the tutor on the one hand, and the one as finally realized by the learner on the other. The simpler a learning route, the simpler its inner representation. In other words, the more complicated it is, the more time its conception, discovery, understanding and recording takes. The simplest possible cognitive network which is repeated at all learning objects demands the least possible concentration on the part of the learner in the course of navigation, therefore emphasis falls on the acquisition of information located in the clusters.

Online learning is, however, not only an individual activity, but its communal, that is collaborative and cooperative nature is also to be highlighted. To this tasks within the course (discussion board, wiki, blog, dynamic glossary, workshop) are necessary to activate students and encourage them to have productive achievements. In this respect Hungarian higher education pedagogy is behind the practice in Western-Europe, the Far-East and North-America.

2. Adaptive E-learning and Web Mining

The distinction between web mining and data mining was made as early as in 1997, however, it only became a field of research in its own right over the last 10 years. There are two approaches to the interpretation of web mining. The process-oriented theory regards web mining the sequence of successive tasks [1], whereas the data-centered concept discusses different web mining methods according to the types of the web data analyzed [2]. It is rather the second approach that has become more accepted, according to which web mining is a special area of data mining, applied for analyzing data created on web servers, that is web content mining, web structure mining, web usage mining are to be discussed. [17] A popular synonym for web mining is the expression 'knowledge discovery in web databases', too. The work of Kosala and Blockeel [3] presents an overview of research into this field up to the year 2000. The study by Srivastava and colleagues also deserves attention, investigating the behavior of users, mostly consumers, at web portals which satisfy great user demands such as Amazon.com, Google, DoubleClick, AOL, eBay, MyYahoo or CiteSeer [4].

From the point of view of our research their endeavours made to identify web metrics and measurements (e.g. visits to pages, visit-purchase rate), to analyze click-streams describing the decision-making process (click-stream analysis) (e.g. the process between entering the web store and purchasing or in fact failing to do so) and to investigate the time factor of web communities, contents and structures are to be highlighted in their work.

Srivastava and colleagues distinguished three phases of web usage mining , that is preprocessing (the identification of users, usually on the basis of IP address; the identification of content by for instance the separation of text, image and multimedia; structure identification as hipertext link), the recognition of patterns in web databases (pattern discovery) (statistical analysis, association rules, clustering, classification, sequential patterns, dependency modeling), as well as analyzing patterns (pattern analysis) (SQL, OLAP) [5].

The application of web mining to examine activities of online learning is a quite special field. Desikan and colleagues used web mining methods for investigating the efficiency of self-directed e-learning [6]. In connection with Desikan's self-directed e-learning we may arrive at three statements with regard to our research. Firstly, the hierarchy of concepts explored in the course of content analysis does not necessarily reflect the appearance of the information on a web page, however, it may still be useful for context and search circumstance analysis. Secondly, the collation of navigation hierarchy created by the planner („expert") and the route actually realized by the learners („beginners") may provide interesting information as to the discovery and comprehension of online learning methods as well as strategies and also to course development. Thirdly, web mining is suitable for modeling the navigation behavior of the learner, too.

In connection with sequence analysis used in our research mention must by all means be made of the five kinds of navigation strategy by Rivers and Storss. They are scanning, that is the overview of a given limited amount of information; browsing, that is the following of a designated route; expedient search; exploration, that is mapping the volume and boundaries of the amount of information; and finally roaming, that is unstructured search [7]. Exploration and scanning may be regarded cognitive strategies.

A virtual learning environment makes all five kinds of navigation strategy possible, however, learners primarily navigate in the system in order to acquire the syllabus. When the designer of the syllabus places diverging clusters in the syllabus he does so in order to provide the learner with liberty in discovering the syllabus. At the same time he would not prefer scanning, let alone roaming to become dominant. In other words, he wishes to maintain the balance between liberty and limitation. In Moodle a certain range of the learning objects may be made visible or even hidden, thus, too, orienting the learning activity of the learners.

Khribi and colleagues still look upon the personalization of virtual learning environment, that is the consideration of the learners' special characteristics both in the course of planning and learning management an as a problem unsolved. That is why they pose the question of an adaptive course management, by which they mean a dynamic restructuring of the course, the adaptive selection and personalized composition of the learning objects as well as an adaptive navigation support. Their adaptive e-learning flow-model can be divided into two phases. In the course of modeling (offline mode) at the formation of the e-learner profile they take into account information gained for example from the interaction between learner and learning environment (e.g. preferred learning objects, learning routes), the existing knowledge of the student, his learning characteristics and style. This is followed by the formation of homogenous groups of e-learners with similar learning characteristics through the application of cluster analysis and associative methods (typifying). In the phase of counseling (online mode) first the observation and analysis of the learning activities of the learner who is just being active in the course takes place along the parameters mentioned in connection with modeling, then the learner is assigned to the group with the same learning characteristics as his. After this propositions concerning learning objects and learning routes are made through the application of filters focusing on syllabus content as well as collaborative activities [8].

A similar model of the formation of a personalized electronic learning environment is to be seen at Jain and colleagues, too, who created an adaptive system by using means of the semantic web as well as web mining. In their model learning objects most suitable for the individual characteristics of learners are selected through the application of so-called personalized e-learning services from standardized content packages, thus creating the so-called personalized electronic learning environment [9].

3. Online Course, Web Mining Methods, Objectives

It was the course implemented in a virtual learning environment (Moodle) of a subject taught in professional teacher training, educational technology and multimedia, that formed the object of the examination. Students taking part in correspondence courses learn the aspects, methods and means of the development of information media (overhead projector foil, video film, photograph, chart, animation, computer presentation, etc) and the use of educational technological aids applicable in the course of their pedagogical work (e.g. overhead projector, video projector, document camera, camera) within the frame of this subject.

An increase in the dynamically changing syllabus content and a decrease in contact lessons necessitated the development and later the application of electronic syllabus in this subject. As a result of the development a four-module (basics of educational technology, digital imaging, image editing, video editing) multimedia based interactive electronic syllabus was created, which, besides the introduction and application in pedagogy of education technological tools, drills the process of the development of information media, that is, the acquisition at a skill and proficiency level of editing programs is highlighted.

Congruence with this dual objective was also reflected in setting the electronic format syllabus content. For the acquisition of information photos, images, texts (written and narrative), animation, and video, while for the introduction of editing algorithms animation supported by narrative explanation and videos were integrated in the electronic syllabus.

Besides the electronic and interactive syllabus contents further objects that support studies were applied in the course, for example discussion boards, wiki and dynamic glossary. Along with that the knowledge of learners was measured through online tests and the solution of productive tasks.

The theoretical model of the educational technology and multimedia course and its concrete interactive syllabus structure were represented in our previous paper [10].

In the course of the examination the CRISP-DM model known in data mining was applied, which is also well usable in analyzing through web mining the database formed in applications for educational purposes [11].

It, however, must be taken into consideration that with respect to a particular course a relatively significant scale, repeated sequence of activities by a relatively small number of test persons is to be taken into account, which in turn makes the generalization of experience difficult.

From among the methods applied during the examination accounts provided by the statistical system of Moodle and in data mining the frequency, sequence and cluster analysis and that of prominent values, classification and associative procedure are to be highlighted.

Data mining procedures appropriate for analyzing online courses and results of their application are treated in several researchers' works [12] [13] [14] [15] [16].

In the course of the research SPSS (IBM) Modeler (Clementine) program and its Web mining node was used, which offers concrete algorithms to realize the data mining methods mentioned above.

For an interpretation of the results of the examination three basic concepts must be made clear. The learner interacts with the screens of the course and its learning objects with the purpose of learning (e.g. opens, downloads and uploads a page or a document, does a test, contributes to the discussion board). So the interaction is to be interpreted at the level of screen pages (php, html, xml), probably files or their larger units, that is events. Therefore analyses focusing on click-streams or learning events (e.g. submitting assignments, activity at forums, wiki-notes) can be distinguished. During a particular visit – from entering the course to leaving it – the learner opens several screen pages, gets into contact with several learning objects and performs a full sequence of operations. The examination directed at click-streams is called microanalysis whereas the one directed at learning events (objects) is called macroanalysis.

All the operations by the learners are administered in the so-called logfile. During the processing, to these notes in the logs User ID, Visit ID and Event ID are assigned, which makes segmentation according to learners, visits and events possible. The date of the operations performed is also recorded, that is tracking the activity of learners in the course is easily solved.

On the basis of the successions of learning activity as well as of the frequency in time of the visits a notion may be formed of the learning methods and strategies. Also, a collation and typifying of visit habits can be realized.

Three variables were introduced for the frequency analysis of the online course. „Days Active" means the time lapse between the learner's first and last visit to the course, „recency" denotes the number of days that passed since the last visit, while learning „frequency" refers to the number of visits during the time interval under examination [10].

Our examination belongs to the category of web usage mining since our objective is the analysis of visit structure, click-streams and learning activity as well as the identification of learner habits, methods and strategies.

Based on the above we looked for an answer to two open questions.

- *Is it possible to draw conclusions of learning characteristics and preferred strategies from the patterns of online learning activities, that is from the order and frequency of student interactions?*
- *How are the syllabus processing procedures as conceived by the developer and as finally realized by the learners relate?*

45 persons (28 men and 17 women), students of a correspondence course in engineering education participated in the research. Their specializations were as follows: 8 mechanical engineers, 16 electric engineers, 6 information technology specialist engineers, 6 light industry engineers and 9 technical managers. With the exception of 4 of them they all teach at vocational secondary schools. The average age of the group is 39.68 years with the youngest person being 27, while the oldest one 58 years old. 16 of them teach in Budapest, 12 in a city and 17 in small towns. 49.17% of the people in the experiment use an IT device in class regularly, 27.69% do so often, while 23.14% rarely. Use mostly refers to a computer presentation, therefore the devices applied are usually portable computers and projectors. It is mainly in the preparation for the class (56.47%) and in communication (76.35%) that IT devices are used outside class. All the persons in the experiment own a computer at home with all but 3 of them with Internet access.

4. Results

The answer to both questions is to be given by an analysis of the learning activities (segmentation of visits by learners or events as well as their temporal analysis).

Our examination may focus on either learning objects (to these learning events such as for example forum activity or submitting assignments) or screen pages (e.g. SCORM or HTML, XML base pages). The former is called macro- while the latter one microanalysis.

Through the examination we arrived at the conclusions below.

4.1 Visit Statistics of the Course

The so-called statistical indexes are efficient supplements to the analysis of learning activities and learning behavior. Fig. 1 shows daily visits, the number of learners, the learning activity as well as the average length of connection. The way the number of daily visits change reflect a weekly periodicity, that is learners on the correspondence course usually study at the weekend. At the beginning of the course they entered the course fewer times but stayed connected longer, while in the second half of the course turnover was bigger but connection shorter. Site Summary Activity Metrics was used for the analysis, however, results were presented with the help of an Excel program.

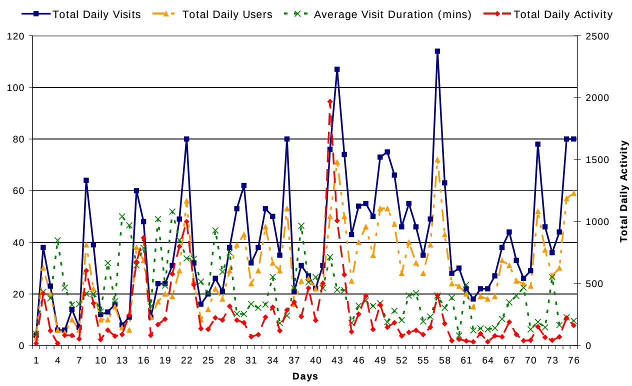


Fig. 1 The learners of the term and visit statistics I.

Algorithm may help answer also the question whether enough time was devoted to processing the given object or not (Fig. 2). To answer this question the contents of the particular pages, as well as the average time necessary for their processing, must be known. Both syllabus units of the basics of educational technology, as well as the first syllabus units of digital imaging and image editing consisted mainly of theoretical skills in the form of textual, visual and audio explanations. The rest of the syllabus units primarily processed editing operations with the help of visual and animation elements and audio explanations. Time devoted to processing the theoretical skills does not appear to be sufficient, which is shown by the average results of the check tests (3.26, 2.97).

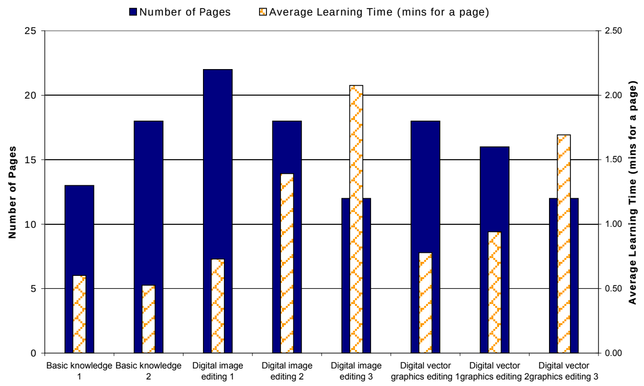


Fig. 2 The learners of the term and visit statistics II.

The Page Usage Metrics stream enables the preparation of statistics segmented according to pages or files. It can for instance be seen over which pages and playing which files (narrative audio, animation, video) learners spent most time. Mediums independent of time head the line with an average of 2 to 6 minutes, while time devoted to visits to screen pages lags far behind with 0.5 to 3 minutes.

4.2 The Segmentation of Learning Activities

An important means of analyzing learning activities is segmentation according to event (learning object), visit (sequence of learning events) or visitor (learner). In both cases we wish to create homogeneous groups from the aspects of learning objectives and learning behavior.

The aim of visit-based segmentation is the identification of the particular patterns and clusters of learning activities, the comprehension of what and why a learner is doing during his activity in the course and finally to aid the developer in recognizing the strengths and weaknesses of the course and in formulating the developmental objectives founded on these.

Based on data characteristics of the use of the Moodle system (time of entry, length and frequency of connection, learning objects used, etc) a learning algorithm groups visits (sequence of learning events) or users (learners in our case). Segmentation takes place in two steps. In the phase of model construction visit segments are created by a so-called two-step clustering, then they are classified by C5.0 algorithm producing decision trees. In the second phase the individual visits are „given scores", then saved in a data file which describes the segments of each visit. The result is the cluster identified, which later may be used for the segmentation of learners, that is the formation of learner groups as well as for certain learning propensity analysis and syllabus-development counseling.

The results of the visit-based segmentation by the Advanced Visit Segmentation stream (focusing on the sequence of learning events) are presented in Table 2. The isolated clusters are as follows:

- *Independent learning directed at the acquisition of basic concepts* which mostly meant the processing of the interactive electronic syllabus of the basics of educational technology, the inspection and extension of the glossary as well as solving the related self-check tests. During their visits to this cluster learners often consulted their teachers with their problems or joined the discussions on the forum. (C1)
- *Learning directed at the acquisition of construction algorithms* which meant the acquisition of the basic operation sequences of editing and imaging programs by viewing or listening to animations and narrative explanations besides the traditional syllabus contents independent of time. Learners did the self-check tests and practice tasks related to the categories (reproductive knowledge). (C2)
- *Participation at moderated discussions* related to subject categories. The objective of these visits was exclusively the inspection or initiation of comments. (C3)
- The upload to the course of *tasks accounting for the creative application of the material learnt* then the inspection of the assessment. (C4)
- *Checking the acquisition of information*, the inspection of the assessment. (C5)

These clusters may also be interpreted as a significant group of the visits focusing on independent learning, while another one on the „justification" of learning achievement (submitting assignments, tests, comments on forums). According to an other interpretation all but one of the activity groups represent the individual form of learning with the one exception representing a community form of it. A mixed pattern of learning objects was less typical.

In Table I the data of each cluster is given. As can be seen the number of visits directed at processing interactive syllabuses (basics, editing algorithms) is relatively small, however, the average time devoted to learning is high, showing significant individual differences though. Visits directed at tests, participation at discussion forums and submitting assignments, in other words the other three clusters, show a picture diametrically opposed to this, that is the number of visits is relatively high, but the length of connection, with the exception of forum discussion, is short and its individual standard deviation lower.

Clusters	Number of visits	Time of connection		Session length (minutes)	Number of activities
C1	207	Morning: 36.71%	M	25.94	23.15
		Evening: 28.5%	SD	34.65	35.35
C2	348	Evening: 34.2%	M	54.86	99.48
		Morning: 26.15%	SD	85.85	87.81
C3	397	Afternoon: 38.85%	M	59.88	75.14
		Evening: 32.12%	SD	70.15	86.45
C4	467	Morning: 52.68%	M	22.43	34.21
		Evening: 24.84%	SD	30.94	30.41
C5	240	Evening: 86.43%	M	17.42	36.02
		Afternoon: 15.62%	SD	14.18	24.02

Session length: the average duration of the visit in minutes; Number of activities: the average of activities per visit

Table I The characteristics of clusters

The creation of homogeneous groups may take place according to learner behavior, too. One means of segmentation is the User Mode Determination stream. Segmentation according to users is similar to the visit-based one. With the use of activity successions the most characteristic learner groups can be isolated by a two-step cluster analysis, from which modes of behavior may be concluded:

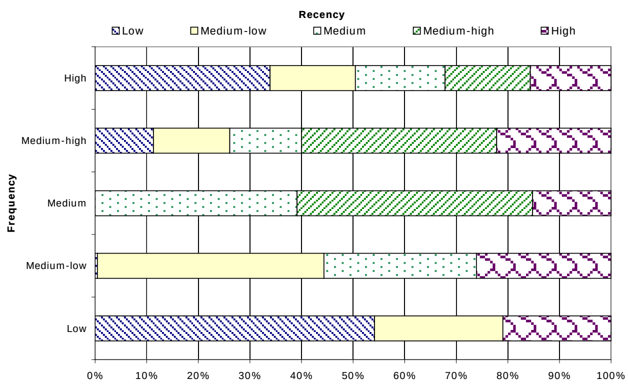
- Interactive syllabus processing by preferred time-dependent media (viewing or listening to animations and narrative audios besides the traditional textual and image contents)
- Interactive syllabus processing by less preferred time-dependent media (ignoring animations and narrative explanations)
- Doing practice tests (related to digital imaging and digital image editing)
- Passive community learning (following forum discussions and wiki notes, reading tutorial messages)
- Active community learning (commenting on discussions, entering notes in wikis and glossaries)
- Checking independent work (doing online tests, uploading productive tasks to the course)
- Assessment of independent work (inspection of the results of online tests and tutorial reflections on assignments)

As we see, all this supports the results of event based segmentation.

As shown by the above, learners present significant individual differences in their course activity. Let us therefore examine the visits in the dimension of time, too.

In order to perform that an other method of segmentation according to learners may help. Segmentation may take place in the dimension of recency, frequency and the individual learning interval (days active). The examination of these variables makes the understanding of online learner behavior easier. For instance learners who have entered the course recently are more likely to return there than those who only visited it some time ago or those learners who frequently visit the course will probably also do so later. The learner who often contributes to a forum discussion will more probably do so later, too.

The E-Channel User RFM Classifications stream is suitable for the identification of online learning behavior in the dimensions of recency – frequency as well as of recency – frequency – days active. Segmentation according to these variables is done by a division into five parts: high (80 percent, top 20%), medium-high, medium, medium-low and low (20 percent, bottom 20%).



$\chi 2 = 692.601$; df= 16

Fig. 3 The segmentation of learner behavior according to frequency and recency

Fig. 3 shows learner behavior in the course with regard to frequency and recency. The rarer learners enter the course the more likely that they made their last visit there a long time ago. Learners of average activity made their last visit only a short time ago. Quite a long time has passed since the last activity of online learners producing a considerable number of visits, in other words these visits were restricted to a short period of time, for example to that before the online check tests or submitting assignments.

Consequently, in this dimension *four types of learner behavior* are to be distinguished.

- The learner visited the course only a few times and briefly and his activity was mostly deadline-related, like for example the submission of some assignment, or an online test, however, it was not preceded by intensive learning on the course (casual online learner).
- Medium-frequency course visits in an even distribution during the term (average online learner).
- The learner visited the course several times but it was always deadline-related, like for example active learning before an online test („campaign learning”).
- A great number of course visits in an even distribution during the term. The learner entered the course several times a week, got informed on forum discussions, viewed a particular syllabus unit, etc (dedicated, loyal online learner).

Average visit frequency is presented in a cross table with regard to days active and recency (Table II). The highest value of days active was 98, the lowest one 64, while the average was 89 days.

		Days active				
		Low	Medium-Low	Medium	Medium-High	High
Recency	Low	0	0	0	0	*0.082*
	Medium-Low	0	0	0	0.029	0.013
	Medium	0	0	0.024	0.013	0
	Medium-High	0	0.018	0.013	0	0
	High	0.019	0.013	0	0	0

The cells contains the mean values of daily visit frequency

Table II The segmentation of learner behavior according to days active and recency

The daily average of visit frequency is highest when days active are on the top and recency the lowest. At the same time, the majority of learners registered in the course or finished their activity by submitting the last task and inspecting the assessment, which in the case of a correspondence course means the end of the term thus leaving the recency value low. It is true for the rest of the cases that days active and recency are in inverse proportion to one another.

4.3 The Analysis of the Syllabus-processing Procedure

Macroanalysis provides great flexibility for the researcher in the identification of event- or object-based learning routes, therefore in the recognition of typical learning habits and strategies.

The interactive syllabus-objects (modules) were divided into smaller, 2 to 3 syllabus units for clarity and achievability, so the particular syllabus units, self-check tests, forums, wikis, etc were classified as independent objects under a given topic (chapter) in Moodle. We were curious to know how learners realized the syllabus-processing procedure as conceived by the designer (syllabus unit 1 – syllabus unit 2 – syllabus unit 3 – self-check test). The Visit Activity Funnels algorithm is suitable for the description of the special patterns of events given in the input data of the stream (learning activity streams) (macroanalysis). During one visit less than 10% of the learners completed a full learning procedure, 35-45% opened the second syllabus unit having completed the first one, and 20-30% processed all the three syllabus units in the course of the same visit. Based on the relation with the particular areas, two kinds of learner behavior are to be distinguished, namely holistic and atomistic. The holistic learner processed all the learning objects belonging to a given area one after the other in the course of the same visit, while the atomistic one had visits directed at one or two objects mostly, in other words, his learning consisted of successions of parts of procedures. Fig. 3 presents visits comprised of the syllabus units of the digital imaging chapter and self-check tests in the course of syllabus-processing.

Item	% Starting Activity	Visits	Dropout Rate
Digital image editing 1	100.00	159	
Digital image editing 2	35.22	56	64.78
Digital image editing 3	17.61	28	50.00
Self-checkTest	8.1	13	53.57

Fig. 4 Learning visits in the area of digital imaging

With respect to forums, wikis and glossaries, two types of behavior can unambiguously be distinguished, namely percipient and productive. The former (55-65% of activities of this kind) opens forum comments, follows discussions, leafs through the glossary, while the latter (35-45% of activities of this kind), besides all that, contributes to discussions and makes comments himself. The number of activities of this kind was low, which is primarily due to features of age. A significant number of corresponding students of engineering education belong to „digital nomads" who mainly „consume" information on the Internet but never or rarely take part in the „production" of it. Special tutorial attention is needed to make them active.

Contrary to macroanalysis, microanalysis focuses on the examination of clickstreams. By this the succession of pages (mostly php and html pages) visited by the learner are meant, which is an efficient means of following learner activity and analyzing the relation to the structure.

Through the analysis of clickstreams the following questions may be answered.

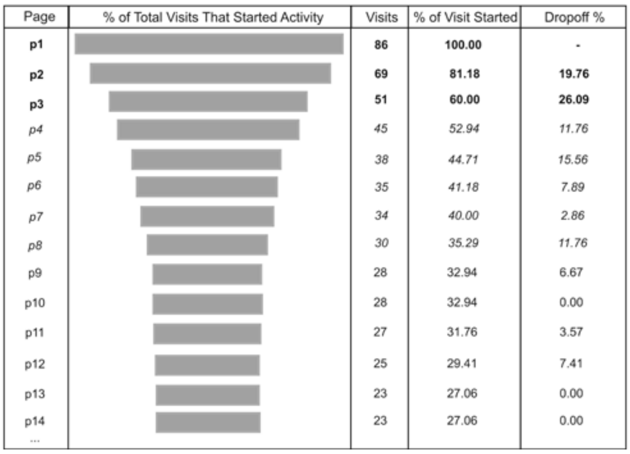
Where learners arrived from an where they went on from the screen page under examination. If the developer's expected route differs from the conceived one, the reason for this may be explored through this.

By analyzing an operation stream between two given pages in a syllabus making bifurcations possible the typical processing routes may be mapped, which may contribute to the identification of certain strategies.

By the representation of page streams between a given minimum and maximum value, related and lasting learning activities are to be identified.

The Visit Page Funnels stream is a means of micro level analysis, looking for clickstream patterns (successive screen pages of syllabus units) defined by the developer among the actual visits realised by the learners. Therefore it proved a suitable means also of comparing syllabus-processing procedures as conceived by the developer and as finally realized by the learners. Three kinds of learner behavior got identified: testing („tasting"), easily giving up and enduring.

The learner who is *testing („tasting"), orientating himself* only takes a look at the first pages, mostly only leafing through them. Those learners who *easily give up* do in fact study, read the textual contents, listen to the narrative explanation, play the animations and videos, however, they do not even get halfway in the syllabus unit. *Enduring learners* succeed to the end of the syllabus unit, not unfrequently turning a page back and forth and mostly thoroughly surveying everything during the very same visit.



Page	% of Total Visits That Started Activity	Visits	% of Visit Started	Dropoff %
p1		86	100.00	-
p2		69	81.18	19.76
p3		51	60.00	26.09
p4		45	52.94	11.76
p5		38	44.71	15.56
p6		35	41.18	7.89
p7		34	40.00	2.86
p8		30	35.29	11.76
p9		28	32.94	6.67
p10		28	32.94	0.00
p11		27	31.76	3.57
p12		25	29.41	7.41
p13		23	27.06	0.00
p14		23	27.06	0.00
...				

Fig. 5 The three characteristic periods of the learning activity

According to Fig. 5 syllabus unit processing all three types are well identifiable. In the course of „tasting" (p1-3) the learner opens the syllabus, reads the objectives and contents, then turns the pages of the document and finally exits. Learners who easily give up do not even reach halfway of the syllabus unit (p4-8) and 10 to 15% interrupts learning. After that, enduring learners usually succeed to the end of the syllabus. These behaviors were mostly characteristic of theoretical syllabuses. In such a case it is advisable for the developer to create shorter units and besides contents independent of time integrate time-dependent media, too, into the syllabus and maintain interest.

Conclusion

Based on the examinations to the open question posed at the beginning of the research, that is concerning online learning specialities, preferred strategies, the following answers may be given:

1. With respect to the relation to the structure

By the application of frequency-, sequence- and cluster analysis four kinds of typical online learning strategies are to be distinguished on the basis of the relation to the structure, namely the conscious and uncertain followers of the structure, structure-abstractor and the unstructured scanner.

The consciousness and uncertainty of structure following is related to the cognitive stylistic characteristics of the field.

The *structure follower online learner* used learning objects mostly in the order given by the course developer (based on the results of the macroanalysis) and in processing the interactive electronic syllabus he consequently followed the route defined by the developer (based on the results of the microanalysis). He primarily focused on the acquisition of the syllabus content and not on getting familiar with the structure and the navigation tools.

The learner independent of the field manages structure with confidence and has mental models ready to find analogy with ease. However, the structure following of the field-dependent learner is characterized by uncertainty, has no mental models ready therefore spends longer time over the processing of a page. A syllabus allowing of multi-bifurcation fits the former type while that of a linear structure primarily fits the latter most.

The *structure-abstractor learner* opened almost all of the learning objects of the course, took the things to be acquired into account and in processing the electronic syllabus contents tried all navigation and bifurcation possibilities. First he „acquired" the structure and only then did he focus on the information to be acquired.

The *unstructured scanner learner* did not visit the learning objects in the order designed by the course developer, his navigation behavior is random-like, often proceeded towards contents which seemed more spectacular and usually only concentrated on tasks with a deadline. He visited the course rarely and at uneven intervals. This learner needs a strictly set syllabus structure and permanent tutorial attention.

The orientation of abstractor and scanner learners is largely aided by the placement of positioning elements in the structure.

2. With respect to the timing of learning

The macroanalysis of learning activities may also be performed according to timing (frequency analysis). For this were learning frequency, period and days active introduced. On this basis four types of learning activity were distinguished: frequent – even (dedicated, loyal online learner), frequent – uneven („campaign learning"), occasional – even (average online learner), rare – uneven (superficial online learner).

Comparing the results of the two kinds of examination we got the patterns seen in Table III.

Although to a different extent, a certain learner „dropping off", the giving up of learning due to cognitive or emotional-volitional reasons, is continuously typical of the processing of syllabus units. Regarding the syllabus units (the result of the macroanalysis) it was mostly typical of the ones processed first, while regarding the screen contents (the result of the microanalysis) it was typical of the first few pages. The abortion of syllabus processing was mostly characteristic of unstructured scanners but also to a certain extent to structure abstractor learners. The latter naturally returned later to continue the syllabus acquisition.

3. With respect to subject areas

Based on the relation to particular subject areas holistic and atomistic learning behaviors are to be distinguished. The *holistic learner* processed all the learning objects belonging to the given area one after the other in the course of the same visit, while his *atomistic colleague*'s visits mostly focused on one or two objects, in other words, his learning composed of a stream of parts of processes.

Learning strategies		The timing of learning			
		Frequent – Even	Frequent – Uneven	Occasional – Even	Rare – Uneven
Relation to the structure of the syllabus	Conscious structure follower	x		x	
	Uncertain structure follower	x	x		
	Structure-abstractor		x		x
	Unstructured scanner			x	x

Table III The patterns of online learning strategies

4. With respect to syllabus- units

According to the length of successive page streams enduring, easily giving up and testing learner behaviors may be distinguished. The first one is characterized by a continuous processing of the entire syllabus unit, the second one by the acquisition of certain parts only, and the third one by inspection which is trying to orientate himself and gather information. While the second one only rarely does so, the third one almost always returns to processing the syllabus unit.

With respect to sensual modality and learning method

Examining learning strategies from the aspect of sensual modality and learning method it can be established that online learners mostly prefer visual syllabus contents and the individual learning method. They generally play digital videos and once the videos have automatically started they are typically never stopped. Narrative audio rarely became an integral part of the online learning process.

Regarding collaborative and cooperative learning methods (active participation in discussion forums and wikis as well as in the extension of glossaries) two kinds of behavior are to be unambiguously distinguished, namely percipient and productive. The percipient learner opens certain forum comments, follows discussions, leafs through the glossary, while his productive colleague himself, besides all that, contributes to discussions, makes comments and is an active participant in wikis.

5. With respect to learner behavior

The modes of behavior identified are largely influenced by for instance the learning characteristics of the individual, their computer attitude, the structure of the course and its learning objects. Taking all this into consideration – with respect to the course examined – , the event-, visit- or visitor-based segmentation of learning activities made it possible to identify variables which have an influence on online learner behaviors: preferred sensual modality (visual – auditive – verbal, and static or dynamic screen contents), mode of comprehension (analytic or global), learning method (individual or social), relation to a social learning space (contemplative or manipulator), the timing of activities (regular, deadline-bound, irregular), relation to the teacher (initiator, reflective, passive), character of syllabus content (informative, for drilling) and relation to structure (structure follower, structure-abstractor, unstructured scanner). Besides all that one must keep the emotional and volitional variables of online learning in mind, too.

References:

[1] O. Etzioni, The World Wide Web: quagmire or gold mine? *Communications of ACM*, Vol. 39 (11), 1996, pp. 65-68.

[2] R. Cooley, B. Mobasher, and J. Srivastava, Web mining: information and pattern discovery on the world wide web. *9th IEEE International Conference on Tools with Artificial Intelligence*, Newport Beach, California, 1997, pp. 558-567.

[3] R. Kosala, H. Blockeel, Web mining research: A survey. *SIGKDD Explorations*, Vol. 2(1), 2000, pp. 1-15.

[4] J. Srivastava, P. Desikan, and V. Kumar, Web Mining - Concepts, Applications and Research Directions. H. Kargupta, A. Joshi, K. Sivakumar, and Y. Yesha (Eds.), *Data Mining: Next Generation Challenges and Future Directions*. American Association for Artificial Intelligence Press, 2004, pp. 51-71.

[5] J. Srivastava, R. Cooley, M. Deshpande, and P-N. Tan, Web Usage Mining: Discovery and Applications of Usage Patterns from Web Data. *SIGKDD Explorations*, Vol. 1(2), 2000, pp. 12–23.

[6] Desikan, C. Delong, K. Beemanapalli, A. Bose and J. Srivastava, Web Mining for Self-Directed E-learning. C. Romero, S. Ventura (Eds.), *Data Mining in E-learning*. (Advances in Management Information), WIT Press, 2006.

[7] Rivers, G. Storss, Characterizing user navigation through complex data structures. *Behaviour and Information Technology*, Vol. 7, 1985, p93-102

[8] M. K. Khribi, M. Jemni, O. Nasraoui, Automatic Recommendations for E-Learning Personalization Based on Web Usage Mining Techniques and Information Retrieval. *Educational Technology & Society*, Vol. 12 (4), 2009, pp. 30–42.

[9] S. Jain, D. K. Jain, H. Bhojak, A. Bhilwar, and J. Mamatha, Personalization of e-Learning Services using Web Mining and Semantic Web. *International Journal of Machine Learning and Computing*, Vol. 2(5), 2012, pp. 569-572.

[10] J. I. Rudas, P. Toth, Examination of Learning Activities by Web Mining Methods. H. Fujita, J. Sasaki (Eds.), *Proceedings of the 9th WSEAS International Conference on Education and Educational Technology*. World Scientific and Engineering Academy and Society, Iwate, Japan, 2010, pp. 31-36.

[11] J. I. Rudas, P. Toth, Web Mining Usage and Quality Assurance. N. E. Mastorakis, V. Mladenov and Z. Bojkovic (Eds.), *Latest Advances in Systems Science and Computational Intelligence*. Proceedings of 11th WSEAS International Conference on System Science and Simulation in Engineering, World Scientific and Engineering Academy and Society, Singapore, 2012, pp. 91-96.

[12] J. Ai, J. Laffey, Web Mining as a Tool for Understanding Online Learning. *MERLOT Journal of Online Learning and Teaching*, Vol. 3 (2), 2007, pp. 160-169.

[13] C. Romero, S. Ventura, and E. García, Data mining in course management systems: Moodle case study and tutorial. *Computers and Education*, Vol. 51(1), 2008, pp. 368-384.

[14] J. Lu, Personalized e-learning material recommender system. *Proceedings of the International Conference on Information Technology for Application*, Utah, USA, 2004, pp. 374–379.

[15] G. J. Hwang, C. L. Hsiao, and C. R. Tseng, A computer-assisted approach to diagnosing student learning problems in science courses. *Journal of Information Science and Engineering*, Vol. 19, 2003, pp. 229–248.

[16] T. Tang, G. McCalla, Smart recommendation for an evolving e-learning system. *International Journal on E-Learning*, Vol. 4(1), 2005, pp. 105–129.

[17] Dzitac, I. Moisil, Advanced AI techniques for web mining. *Proceedings of the 10th WSEAS International Conference on Mathematical methods, Computational techniques and Intelligent systems*, 2008, Wisconsin, pp. 343-346.

[18] Toth, R. Osz, A Study of Online Learning by Methods of Web Mining. *Proceedings of 12th WSEAS International Conference on Education and Educational Technology*, World Scientific and Engineering Academy and Society, Singapore, 2013, to be published