Conducting a Web Browsing Behaviour Study – An Educational Scenario

Martin Labaj, Mária Bieliková

Slovak University of Technology in Bratislava,
Faculty of Informatics and Information Technologies,
Ilkovičova 2, 842 16 Bratislava, Slovakia
{martin.labaj, maria.bielikova}@stuba.sk

Abstract. Web browsing behaviour is a matter of study in several fields – from web usage mining, to its applications in adaptive and personalized systems. Current web browsers allow for parallel browsing – opening multiple web pages at once and switching between them. To capture such behaviour, client-side observations are typically performed, where attracting and retaining enough participants poses a challenge. In this paper, we describe a study based on an experiment on logging the parallel browsing behaviour, both in an adaptive web-based educational system and on the open Web, while using the educational system as a tool for recruiting and motivating the participants. We focus on how various types of users (here students), including their personality information, participated in the experiment regarding churn and their observed behaviour. The paper concludes with "lessons learned" important to consider when planning and performing similar studies.

Keywords: web browsing study, tabbed web browsing, educational systems, churn

1 Introduction and Related Work

As more and more aspects of everyday life are now being carried out using the Web, if not completely on the Web, observing and analysing web browsing behaviour comes into importance. By understanding how users browse websites, either within a single web-based system, or across heterogeneous systems on the open Web in general, we can possibly improve any personalization and adaptation feature that can be based on user behaviour (i.e. implicit feedback), such as:

- Domain modelling – creating links between pages, links between domain terms, or assigning domain terms to content based on user movement across the Web.
- User modelling – discovering user's current interests and predicting future ones from their visits to pages.
- Recommender systems – observing patterns in users' visits from a page to other pages and recommending relevant pages to similar users.

An important aspect of how users browse the Web stems from modern web browsers. These not only allow opening multiple windows at once, but also allow opening multiple web pages within a single window using tabs. Such behaviour is called parallel browsing [1, 8] or tabbed browsing [4, 9, 18]. Since multiple web pages are accessible to the user without having to load or reload them, the tabbing has changed the traditional web usage mining approaches [18].

Traditionally, server-side log-based web usage mining considers *page load* events, but with the parallel browsing, the user can switch contexts by navigating between opened tabs with various pages attributed to different user tasks without generating these page load events that could be visible to server-side logging. The user can also open a page in various ways. A link can be opened into the same tab, replacing the currently opened page, or in a new tab, leaving the source page opened and branching the browsing action tree. Therefore, even when the page load event is observed in the server-side logging, it can represent differing browsing actions when we take parallel browsing into account. Various inferences can be made from traditionally logged actions, e.g. when the user loads two different pages by clicking a link from the same page, the page must have still existed for the second click and therefore the first click must have been branching [7, 18]; or tasks in the browsing can be discovered and then browsing behaviour can be estimated [1]. However, in order to *fully* capture and analyse the user browsing behaviour including using multiple pages (tabs) and switching between them, and to allow for the aforementioned personalization and adaptation features based on user behaviour, one must use *client-side* tracking and observe the actions on the user computers in their browsers.

A single web-based system can easily track its users on the client-side by including client-side-executed scripts into pages served by this system and aggregating user actions across multiple pages (activating/deactivating single pages) to reconstruct the browsing tree. We previously used such approach [10] to capture user browsing within adaptive web-based educational system ALEF (Adaptive Learning Framework).

Users, however, do browse in tabs across heterogeneous web systems in various situations [4], ranging from comparing pages against each other, keeping frequently used ones at hand, creating bookmarks or todo lists, to simply multitasking. To observe such behaviour, the tracking must be done either in all pages, e.g. by using an adaptive proxy that injects scripts into web pages passing through it, or by observing the user's browser directly via a browser extension. Except when monitoring connections in a school, in a workplace, etc., where an intercepting proxy [12] or browser can be configured authoritatively, active user participation is required to install modified network settings or a browser extension [4] to participate in such experiments.

Studies of browsing behaviour therefore often face a choice, either:

- *Passively use server-side* data that are easily observable from all visitors, but those do not provide details on advanced features of their browsing behaviour, such as parallel browsing. Or:
- *Actively monitor the participants with client-side software*, but attracting enough participants to install the logging software voluntarily (in order to observe their natural behaviour) and retaining these participants usually becomes a challenge.

In this paper, we describe a study where we observed users' browsing behaviour both within adaptive web-based educational system ALEF (Adaptive Learning Framework) [17] and on the open Web while using ALEF as a tool to find participants and motivate them to stay in the study, i.e. to observe the web browsing behaviour to maximum extent, while actually reliably finding and retaining participants. We focus on how students participated in the experiment.

Our aim is to understand how users participate in a voluntary long-term browsing behaviour study that requires installation of a logging browser extension which is, in fact, always a privacy intrusion to some degree, and how long they remain participating. By observing user features that can be observed independently from the study, such as participant demographics elicited through surveys, their academic performance (in the case of student users), or their actions in other web systems (such as in an adaptive learning system), we could predict how would users join before even starting the study and for example estimate required population to be acquired. If we could predict how long would they participate (their churn rate) and what causes them to stay/leave, we could obtain more results from more participants in similar studies.

The rest of this paper is organized as follows: first, we describe the browsing behaviour logging infrastructure and explain the experiment setup, including motivation and recruitment of the participants via ALEF. Next, the resulting dataset is described, followed by analysis of churn and user behaviour during the experiment. In the last part, we outline future work and implications of this study.

2 Study on Logging the Browsing Behaviour

2.1 Experiment Setup

For the purpose of this experiment, we implemented parallel browsing tracking as an extension in the Brumo platform (Browser-based User Modelling and Personalization Framework[1]) [15], which provides infrastructure for browser extensions, including client-server communication and storage, and allows distributing the extensions. We capture user actions in separate tabs, such as loading the page, bringing the page into focus or hiding it, and then combine these actions into single-timeline stream using reconstruction algorithm we described in [10]. We created an infrastructure, where the events are observed using the extension, sent to server and analysed [11]. The output is a browsing tree describing how a given user clicked on each single link or typed-in URLs (into the same tab or a new tab), how they switched between tabs or out of the browser – effectively reconstructing entire user session, allowing for analysis of switch frequencies between given resources, or user browsing styles.

We set up experiment with participants who used educational system ALEF – bachelor students of the Principles of software engineering course – instructing them to broaden their knowledge about topics presented in the system and look for appropriate external sources (URLs related to given content) on the open Web, attaching them to corresponding learning objects. This task was also motivated competitively

[1] http://brumo.fiit.stuba.sk

using global user score and leaderboards – students were awarded score points for links attached to the content [6].

Submitted external sources were scored according to their novelty (repeated URLs were penalized), access level (sources attached as public and signed with own username were worth more points than anonymous sources), and finally according to their quality and relevance to the given learning object. The last criterion was evaluated by a domain expert, who rated the sources in three levels: appropriate (accepted), neutral, rejected, which were rewarded with more points, rewarded with the default amount of points, or penalized with negative amount of points, respectively. On top of that, whenever a student achieved a reward level amounting to five top-rated (signed and approved) sources, an additional content (a recapitulating question for a final exam) was unlocked for the given student – motivating the students to browse for quality links.

Most importantly, we asked the students to let us see their browsing behaviour while looking for these external sources and in order to do that, the students could insert the external sources only when the Brumo extension with browser tracking was present. The main phase of the experiment (inserting the external sources) ran for more than a week and data collection continued voluntarily for a year among users who have kept the extension installed after the learning system experiment.

2.2 Dataset

For studying various user approaches in browsing style and participation in the study, we created a dataset consisting of 249 users. The users are structured into three groups:

1. 80 users: bachelor students from ALEF system who chose to participate in the browsing study. These are students who were learning the course content in ALEF and as other ALEF students, they were given the opportunity to install the extension and then submit external sources that rewarded points as a motivation. These students chose to install the extension in order to participate. Then they could leave the extension installed for some time, remaining in the study. This group serves as a *positive* sample and allows exploring which features influences joining the study and the length of participation.
2. 144 users: bachelor students from ALEF system who chose not to participate in the browsing study. These students were also exposed to the same motivation as the group (1), but chose not to participate. However, except the browsing behaviour on the open Web (which is naturally not available, since these students did not install the logging extension), other features about their activity in the learning system, personality, etc. (see below) were still observed for these users, therefore this group serves as a *negative* example and allows exploring which student/user properties have impact on not participating.
3. 25 users: older students (mostly master study) in the browsing study who did not come from the ALEF system. These users started using the logging extension through other means, for example by attending a student research seminar where

the extension was propagated. The number of these users is rather low, because the logging extension was freshly deployed and there are little additional features known about these users apart from their browsing behaviour.

Additional to browsing actions within and outside of the educational system, we computed/included churn data (when the user joined, left, etc.), activity within the experiment (external sources statistics), activity within the educational system outside of the experiment (question-answer learning object answering), and study performance. Two types of demographic traits were elicited: Felder and Silverman learning style [5] was obtained via a questionnaire within ALEF, and Big Five personality traits were obtained via professional assessment. Table 1 provides more detail about the data collected in the dataset.

Since several of these attributes depend on participation, either by doing some voluntary activity in the educational system, or filling out the questionnaire, or taking the personality traits assessment, not all of these data are available for each user. Dataset features coverage is shown in Figure 1.

Table 1. Dataset composition.

Feature		Source	Explanation
A	Churn	Brumo extension, partially inferred	Date and time joined and left (started/stopped using the extension), participated in experiment, is still active
B	Browsing actions	Brumo extension, ALEF	Browsing behaviour within ALEF, outside of ALEF, during experiment, outside of experiment, total browsing
C	External sources	ALEF	External sources submitted, categorized as approved, rejected, and deleted
D	Learning activty	ALEF	Number of shown and number of answered question-answer learning objects; portion of views skipped without answering (tendency to “cheat” into viewing question-answer pairs without having to rate); portion of views rated with default value (similar)
E	Study	Course	Academic performance
F	Learning style	Questionnaire (ALEF)	Dimensions (active/reflective, sensing/intuitive, verbal/visual, sequential/global)
G	Personality traits	Professional assessment	Traits (openness to experience, conscientiousness, extraversion, agreeableness, neuroticism) – value and percentile

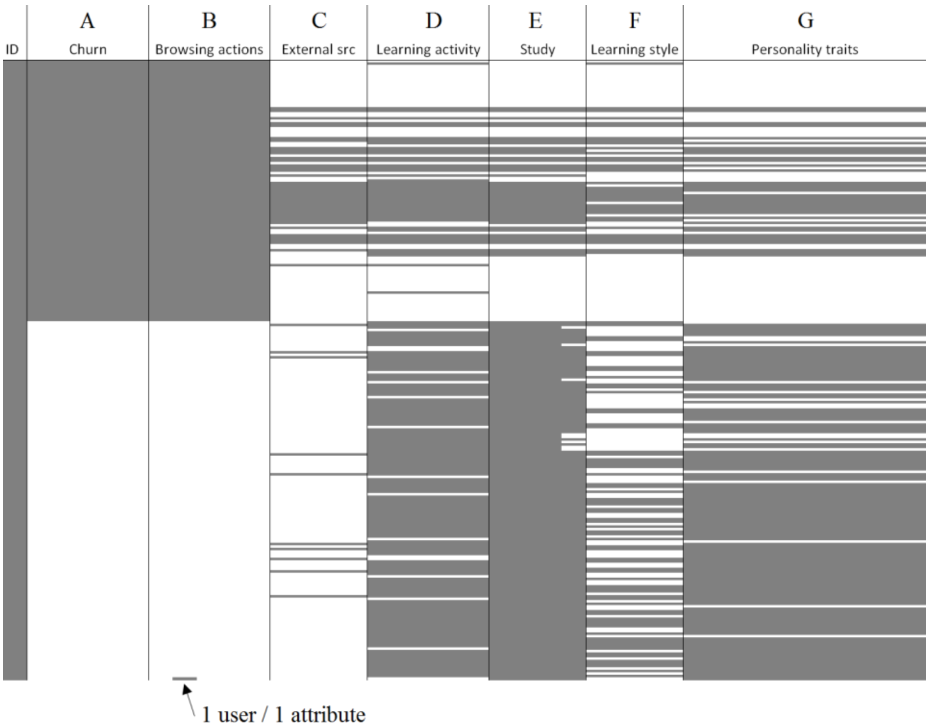

Fig. 1. Dataset sparsity: first column represents participations (by their ID), columns 2 to 8 represent particular features (according to Table 1); grey – attribute is present, white – attribute is missing (user did not participate in the source activity). First 105 users participated in the browsing study (groups 1+3), rest of them, 144, did not participate (group 2).

2.3 User Participation and Browsing Style

First, due to the nature of the continuing experiment, we were interested in churn rate (or retention conversely). Determining the churn is a task of finding when the user is likely to leave. The term comes from the telecommunications field, where one is interested in when and why is the user going to switch to competing service provider, but the churn is commonly explored in the field of adaptive web-based systems. For example, in a user-generated content community, rate of participation and feedback from other users can be associated with length of membership [14]. In another example, a length of participation in a community-based question answering system can be predicted using classifiers based on questions, answers, gratification and answerer demographic features [3]. Perhaps closer to our study is an example of content discovery system, where users view recommended Web pages. Time spent, visit features and content features can be associated with number of sessions made by the user [2].

If we could predict the churn, we can, for example, increase or add user motivation during the experiment to prevent them leaving and obtain better data, or even predict how many participants we need in the beginning given the expected dropout through the course of experiment. We believe that simple actions like notification of the teacher or stimulated message to the student including gamification measures currently more and more employed in domain of educational systems can further improve students' engagement and so their results. There is, however, a rather important difference to traditional churn prediction tasks. In setups such as ours, the user is being observed in their natural activity on the Web and it is even desirable for the purpose of obtaining unbiased data to interfere with this activity as little as possible. Actions on the web therefore do not correspond with the user satisfaction or engagement in the experiment. In our case, we have multiple external features relevant to the user – study performance in the course, activity in ALEF, external source activity, or personal traits which could be possibly used universally.

A potential candidate for relevance to churn is the activity in ALEF related to question-answer learning objects, because the motivation is similar [16]. The user is presented with a pair consisting of a question and a student-provided answer and has to judge the correctness of the answer. The user is motivated by getting acquainted with potential questions for upcoming exams and recapitulating own knowledge. In the external sources gathering experiment, the students were rewarded for multiple quality links with a potential exam questions, so there is similar motivation in both of these activities.

Users can, however, "cheat" out of the work of rating the answer correctness by using an option to skip a question and get another one multiple times in a row, or by answering with the default correctness value. While some rate of skips and default answers is natural, for some users, rates as high as 90 % skipped questions were observed. This could relate to the user's approach to participation in other kinds of activity – e.g., if a user is only interested in the score and reward levels for the external sources, he/she could install the logging extension for a very brief period of time or even into a separate browser, submit prepared sources (to achieve a score for the attached sources) and stop participating.

The overview of how users joined and left the study is shown in Figure 2. The overview of relations found in the dataset is shown in Figure 3. Correlations within segments of features (group of features describing on property of subjects, e.g. their learning style) delimited by horizontal and vertical lines are expected, since some of the features are inferred from others in the same segment, or they may be related on each other, for example, the ALEF experiment ran from the beginning of the dataset for some time, therefore time of joining is related to the user having participated in the study from the ALEF experiment (users who joined sooner), or having participated in the study independently (users who joined later).

The question-answer learning object activity described above correlates both to whether the user has participated in the study at all (0.297, p-value = 0.002) and how actively they participated (0.483, p-value $\approx$ 0), suggesting that these activities with similar underlying motivations (allowing the students to practice exam-like questions) attracted similar user behaviour. Therefore, *if we base user motivation in an experiment starting from an educational system on similar mechanics as another activity in the system with known usage, even when the activity is very different to the experiment, we can predict user participation to a degree*.

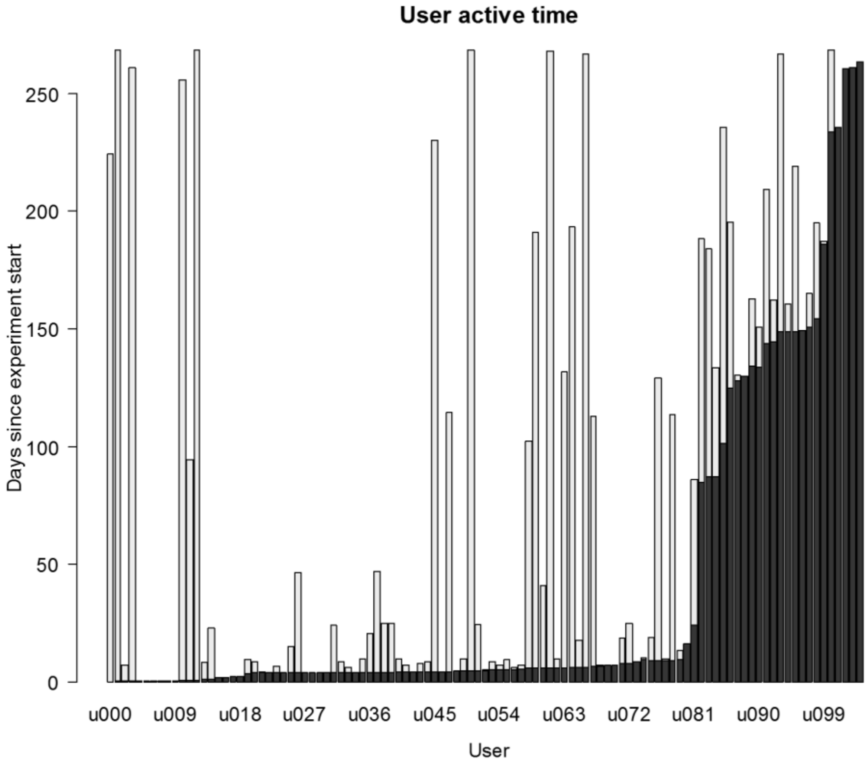

Fig. 2. User participation in the browsing study (groups 1 and 3). Dark bars represents time (days) since the start of the study until the user has joined. Light bars are the length of the participation (time until the user has left).

In the logs themselves, we indeed found a behaviour we can call "*short-participating*". It is similar to the skipping and default-rating the question-answer learning objects – some users allowed tracking only for the minimal time, when they installed or enabled the tracking, looked up and attached several external sources to the learning content (inserting the sources into the system was the only activity that strictly required the presence of tracking) and stopped the tracking altogether. In some cases, only the inserting took place with the tracking. This can be induced by realising that the tracking is needed to insert the link only after having the external links already found, in that case, clearer communication (within the constraint of this being an uncontrolled experiment) may help.

In some cases, students mentioned that they use a different browser for their daily browsing activity and they used the two browsers for which the tracking is implemented only to insert sources looked up elsewhere. Therefore, in browsing behaviour

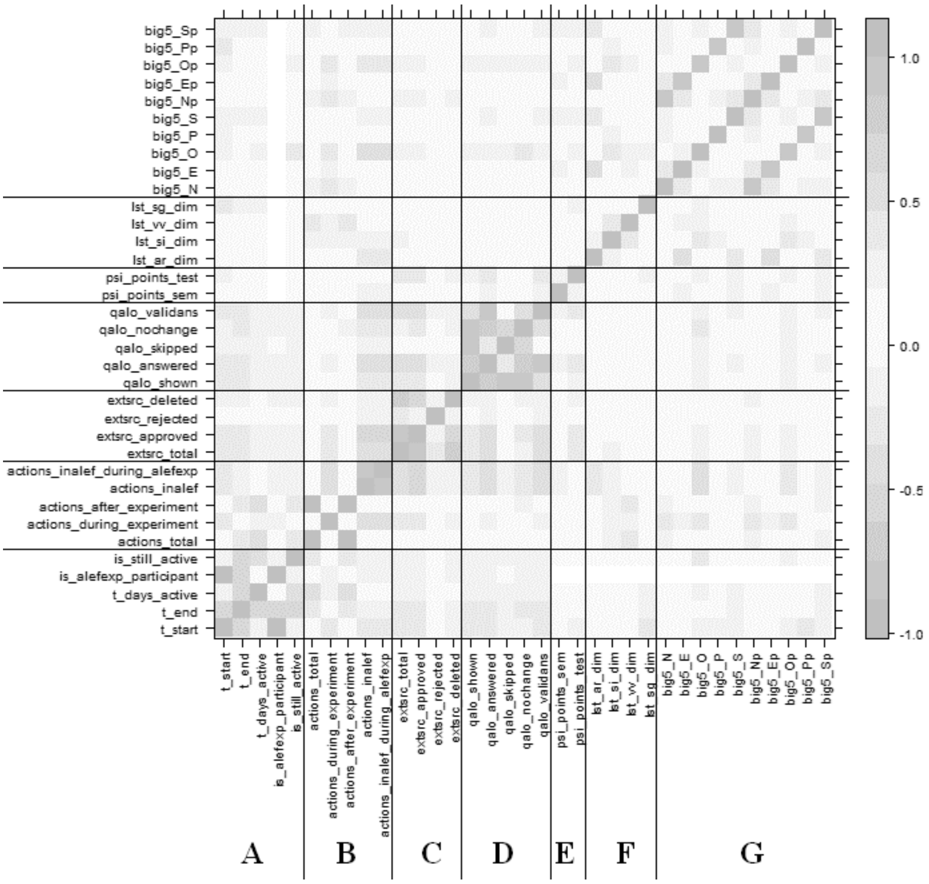

Fig. 3. Associations between various features of users found in the experiment (Pearson pair-wise correlation). Horizontal and vertical lines mark feature segments of the same type. The notation identifies features according to their segment explanations, e.g., personality dimensions, learning style dimensions, or browsing actions (see Table 1).

studies, it is important to *cover as many various browsers as feasible technically and labour-wise*. Due to differences between browsers in creating extensions, integrating with the browser and pages, maintaining their compatibility through new browser versions, etc., covering every possible browser is infeasible. Often only one browser is observed.

If a study is concerned with a smaller number of participants (such as when follow-up in-person interviews are planned), they can be recruited with explicit requirement of everyday usage of a given supported browser, such as in [4]. However, in cases similar to our where we motivated users to participate, temporary switching the browser can hinder the goal of observing the natural user behaviour on a large scale. Note that we supported two common browsers (Chrome and Firefox), such as some other studies do [19], and yet there was a relatively significant portion of users who commented on using other browsers (Safari, Opera). On the other hand, supporting two and more browsers creates a possibility for users to intentionally install the extension into that browser they do not use naturally, cases of this behaviour were observed in this study.

We then looked on the *continuing participation* in the study. On one hand, using a "stealth" extension (one that does not present any specific features directly to the user, although it is not hidden in any way in the standard extension-listing user interfaces in the browser) to track browsing has an effect of observing the natural behaviour of the users and some participants will continue to use it because it does not interfere with their browser experience. On the other hand, this very same mechanism of the initial motivation (e.g. user score, questions) in an experiment and no subsequent user features of the tracking can mean that we lose participants, e.g., when the user reinstalls or switches the browsers. In interviews with selected participants, several of those who left the study at some time after the experiment (they did not short-participate as above) mentioned that no user features were the reason for deliberately uninstalling the tracking.

It seems that *the best approach is combining the initial motivation (as described here with ALEF experiment) to gather the initial user base together with later features available in the tracking software (allowing the users [of educational system] to see their browsing history, manage tabs, etc.) to retain the users for ongoing data collection.* Instead of having the logging stay quietly behind, making the logging software attractive to participants can not only help keep users in the study, but possibly also attract new users who did not receive or did not respond to the initial motivation such as that from the educational system we describe here.

Regarding the browsing behaviour within the adaptive learning system, a common type of browsing paths observed in the experiment was a "*loop*", which is a sequence of tabbing actions that:

- starts in a learning object,
- follows user switches through several pages and
- returns to the same learning object.

Such feature could be, for example, used for discovering external sources complementing the content that the user was previously browsing. A method for web content enrichment was created based on the logged browsing behaviour [13].

3 Conclusions and Future Work

In this paper, we presented findings and *lessons learned* of conducting a study on open Web browsing behaviour of students who use an educational system, which served as a motivation tool for engagement in the study. Our study was based on the dataset consisting of 249 users structured into three groups according to activities they performed within the educational system ALEF. We explored user participation and found a correlation with participation in the system activity. We summed up our observations as recommendations for conducting browsing studies. These include using other activity in an adaptive web based system (here educational system) as a basis for predicting the participation in another study, avoiding “short-participating” when covering single, or on the other hand, multiple web browsers, and continuing the participation motivation in some form after the initial drafting motivation. By caring for these aspects, a browsing study can gather more participants, observe as much as possible of their natural behaviour (both in terms of length of the observed part and in its quality).

In spite of the fact that the resulting dataset is based on enough participants to consider their behaviour and traits, the number is insufficient for cross-validation of a predictive model. It, however, includes diverse set of attributes describing various aspects of the user from personality traits, to performance, to activity in an educational system, and activity on the open Web, which helps to get inside into browsing behaviour of study participants. Along the lines of such breadth oriented study, a qualitative experiment observing physical users could complement these findings.

Research described here was so far focused on the user side of the problem, i.e. how users participated and browsed. Another aspect, the item based view, on which we plan focusing now, is important for leveraging these findings in domain modelling, recommender systems, etc. – how browsing styles differ depending on the item. We have shown previously that the tabbing actions differ in learning environment based on the type of learning object (explanations, exercises, and questions) [10]. Various browsing styles could perhaps differentiate the quality of web pages, or help us reveal relations between objects, such as subsumed-by (one object logically follows after another, especially in, but not limited to, an educational environment), or related-to (one object explains concepts present in another object).

Acknowledgement. This work was partially supported by grants No. VG 1/0646/15, KEGA 009STU-4/2014 and it is the partial result of the Research and Development Operational Programme for the project “University Science Park of STU Bratislava”, ITMS 26240220084, co-funded by the European Regional Development Fund.

References

1. Bonnin, G., Brun, A., Boyer, A.: Towards Tabbing Aware Recommendations. Proc. of the First Int. Conf. on Intelligent Interactive Technologies and Multimedia - IITM ’10. pp. 316–323 ACM Press, New York, USA (2010).
2. Dave, K.S., Vaingankar, V., Kolar, S., Varma, V.: Timespent Based Models for Predicting User Retention. WWW ’13 Proc. of the 22nd int. conf. on World Wide Web. pp. 331–342 (2013).
3. Dror, G., Pelleg, D., Rokhlenko, O., Szpektor, I.: Churn Prediction in New Users of Yahoo! Answers. Proc. of the 21st int. conf. companion on World Wide Web - WWW ’12 Companion. pp. 829–834 ACM Press, New York, USA (2012).
4. Dubroy, P., Balakrishnan, R.: A Study of Tabbed Browsing Among Mozilla Firefox Users. Proc. of the 28th int. conf. on Human factors in computing systems - CHI ’10. pp. 673–682 ACM Press, New York, USA (2010).
5. Felder, R., Silverman, L.: Learning and Teaching Styles In Engineering Education Richard. Eng. Educ. 78, 7, 674–681 (1988).

6. Filipčík, R., Bieliková, M.: Motivating Learners in Adaptive Educational System by Dynamic Score and Personalized Activity Stream. Proc. of 9th Int. Workshop on Semantic and Social Media Adaptation and Personalization – SMAP'14. IEEE. To appear.
7. Huang, J., Lin, T., White, R.W.: No Search Result Left Behind: Branching Behavior with Browser Tabs. Proc. of the 5th ACM int. conf. on Web search and data mining - WSDM '12. pp. 203–212 ACM Press, New York, USA (2012).
8. Huang, J., White, R.W.: Parallel browsing behavior on the web. Proc. of the 21st Int. Conf. on Hypertext and Hypermedia - HT '10. pp. 13–17 ACM Press, New York, USA (2010).
9. Chierichetti, F., Kumar, R., Tomkins, A.: Stochastic Models for Tabbed Browsing. Proceedings of the 19th int. conf. on World wide web - WWW '10. pp. 241–250 ACM Press, New York, USA (2010).
10. Labaj, M., Bieliková, M.: Modeling parallel web browsing behavior for web-based educational systems. Proc. of 10th Int. Conf. on Emerging eLearning Technologies and Applications – ICETA'12. pp. 229–234 IEEE (2012).
11. Labaj, M., Bieliková, M.: Tabbed Browsing Behavior as a Source for User Modeling. Proc. of User Modeling, Adaptation, and Personalization. pp. 388–391 Springer (2013).
12. Meiss, M., Duncan, J., Gonçalves, B., Ramasco, J. J., Menczer, F.: What's in a Session: Tracking Individual Behavior on theWeb. Proc. of the 20th conf. on Hypertext and Hypermedia - HT '09. p. 173 ACM Press, New York, USA (2009).
13. Račko, M.: Automatic Web Content Enrichment Using Parallel Web Browsing. Proc. of 10th Student Reasearch Conference in Informatics and Information Technologies - IIT.SRC'14. STU Bratislava, pp. 173-178 (2014).
14. Sarkar, C., Wohn, D.Y., Lampe, C.: Predicting Length of Membership in Online Community "Everything2" Using Feedback. Proc. of the conf. on Computer Supported Cooperative Work Companion - CSCW '12. pp. 207–210 ACM Press New York, USA (2012).
15. Šajgalík, M., Barla, M., Bieliková, M.: Efficient Representation of the Lifelong Web Browsing User Characteristics. Proc. of the 2nd Workshop on LifeLong User Modelling, in conjunction with UMAP 2013. pp. 21–30 CEUR-WS (2013).
16. Šimko, J., Šimko, M., Bieliková, M., Ševcech, J., Burger, R.: Classsourcing: Crowd-Based Validation of Question-Answer Learning Objects. Computational Collective Intelligence. Technologies and Applications. Proc. of 5th Int. Conf., ICCCI 2013, pp. 62–71 Springer (2013).
17. Šimko, M., Barla, M., Bieliková, M.: ALEF: A framework for Adaptive Web-Based learning 2.0. Key Competencies in the Knowledge Society, WCC 2010. pp. 367–378 Springer (2010).
18. Viermetz, M., Stolz, C., Gedov, V., Skubacz, M.: Relevance and Impact of Tabbed Browsing Behavior on Web Usage Mining. Proc. of IEEE/WIC/ACM Int. Conf. on Web Intelligence - WI 2006,. pp. 262–269 IEEE (2006).
19. Von der Weth, C., Hauswirth, M.: DOBBS: Towards a Comprehensive Dataset to Study the Browsing Behavior of Online Users. Proc. of the 2013 IEEE/WIC/ACM Int. Joint Conf. on Web Intelligence (WI) and Intelligent Agent Technologies (IAT) - WI-AT '13. pp. 51-56 IEEE (2013).