Behavior Based Adaptive Navigation Support

Michal Holub
Institute of Informatics and Software Engineering
Faculty of Informatics and Information Technologies
Slovak University of Technology
Ilkovičova 3, 842 16 Bratislava, Slovakia
holub@fiit.stuba.sk

Mária Bieliková
Institute of Informatics and Software Engineering
Faculty of Informatics and Information Technologies
Slovak University of Technology
Ilkovičova 3, 842 16 Bratislava, Slovakia
bielik@fiit.stuba.sk

ABSTRACT
Web portals contain large amount of information. Users could really benefit from it if personalized presentation is used. For this to be accomplished the website needs to “know” its users. When surfing the Web users leave digital footprints in the form of navigational paths and actions taken. We present a method for adaptive navigation support and link recommendation. The method is based on an analysis of the user’s navigational patterns and behavior on the web pages while browsing through a web portal. We extract interesting information from a web portal and then recommend it. Finally, we provide our experience with a recommender system deployed on our faculty’s website, which recommends events by means of personalized calendar.

Categories and Subject Descriptors
H.5.4 [**Information Interfaces and Presentation (e.g., HCI)**]: Hypertext/Hypermedia—*Navigation*. H.3.3 [**Information Storage and Retrieval**]: Information Search and Retrieval—*Relevance Feedback*.

General Terms
Algorithms, Design, Experimentation, Verification.

Keywords
Adaptive navigation support, automatic interest estimation, behavior, link recommendation, navigational patterns.

1. INTRODUCTION
Web portals are being visited by various users pursuing different goals. However, most websites offer all groups of visitors the same content. Therefore the visitors are often presented information in which they have no interest [4].

While browsing through a web portal some users can discover interesting pages that are hidden deeper in the hierarchy of the portal. If the users with similar goals knew about these pages they could find them interesting, too. Personalized navigation and recommendation based on monitoring user activities and social principles is a viable approach for such cases. In this paper we introduce a method of navigation adaptation and social recommendation of links among users with similar behavior.

2. RELATED WORK
People use different means of accessing information on a web portal. The most common way is to follow hyperlinks, which accounts for more than half of all possibilities [7]. This introduces a problem with improper navigation containing large number of links. The user has often a problem of deciding which link to use. Therefore a recommendation of links on the web page could bring significant improvement to user’s browsing experience. Other dominant mean of navigation is using the browser’s back button [12]. Accessing websites through the history, bookmarks and other means is insignificant.

User’s habits can be derived from the navigational patterns found in the sequences of links he uses in a particular web portal. Four basic navigational patterns (path, loop, ring and spike) were described in [6]. From the prevailing patterns in browsing sessions different browsing strategies can be identified.

User’s interests are often determined based on the content of documents the user has read [5]. The user model can be expressed by concepts or keywords extracted from these documents [2]. If we know what topics (usually expressed by the keywords) the user prefers, we can recommend him documents (web pages) with similar content. The disadvantage is that documents should be written in language which we can process (a translation can help).

In [16] authors use rather different approach based on user behavior tracking to estimate his interest. For this we need to get a feedback from the user. There are several ways how to implicitly determine user’s interest. When links are well annotated (like on news portals where links to articles contain a short introduction) the event of clicking on the link is considered as positive interest [8, 14]. However, in general scenarios we cannot always consider clicking on a link as truly positive interest in the web page.

To determine user’s interest we can also use actions he makes on a web page [11]. Printing the page or adding it to bookmarks show positive interest. Spending very short time reading it or even closing the browser prematurely shows negative interest [16].

With user’s interest determined navigation personalization as well as link recommendation can be done [9, 13]. In [10] authors propose a method of interesting link recommendation by highlighting the links. They extract keywords from the pages a user visits and recommend links that lead to other pages which contain the same keywords. Adaptive system *Web Watcher*, which implements this approach, can also show similar pages to the page that is currently being viewed based on this principle. The system uses a proxy server to incorporate its toolbar into every web page.

Other method is based on monitoring the context in which the links are being used [1]. This method consists of creating a knowledge base from the links each user has clicked on. Then the clusters of links, which are often used together, are built from the knowledge base. Links from a cluster with the largest overlap with the current session of the user are recommended to him.

All methods mentioned share the same feature which is user interest estimation based on his actions. They prefer behavior of the users over content of the documents which they were shown.

3. ADAPTIVE LINK RECOMMENDATION

We propose a method for adaptive recommendation of interesting links in a particular web portal (which we may or may not own). For a specific user we select links that similar users found interesting. We also recommend links to this user based on his previous surfing sessions. Our recommender system extracts further information from the web pages, which is also shown to the user. The recommended links are shown in special sections added to each web page of the portal.

When deciding which link to recommend we do not consider the content of web pages. We based our recommendations solely on the analysis of user's behavior. Our method thus does not depend on the language of the website. We are able to analyze interest and patterns on different language versions of the same portal. Our method of adaptive link recommendation works in two steps:

1. Mining web usage history for navigational patterns.
2. Recommendation of links based on user's behavior.

In the first step we analyze the sequences of followed links from each user's session. In these sequences we look for navigational patterns. We use the prevailing pattern to categorize users, as it determines user's surfing habits on a particular web portal. As an output we get groups of users with similar navigational patterns.

In the second step we monitor behavior of users on each visited web page of the portal. From their actions we automatically estimate their interest in that page. We then recommend links to interesting pages among users of each group from the step one.

3.1 Discovering navigational patterns

We find similar users based on comparison of navigational patterns they follow in a closed web portal. We believe that users who follow analogous paths should be recommended similar links. There are four basic navigational patterns according to [6]:

- *Path* – a sequence in which nodes do not repeat.
- *Ring* – a sequence that starts and ends in the same node.
- *Loop* – a sequence that goes through already visited node.
- *Spike* – a sequence that goes back through the same trail.

In each session a user visits several pages of the web portal. This session is described by a vector whose elements are links to the web pages arranged in order they were visited. We consider a continuous sequence of links to be a session. For this purpose we use the *referrer* field of HTTP request message. If the URL of previously visited page equals referrer value of currently visited page, we consider the pages to be in the same session.

The process of dividing users into groups is presented in Alg. 1. Similarity of users is expressed as Pearson correlation coefficient [15] commonly used in collaborative filtering.

Algorithm 1 Group users according to their similarity.

```
for each user u do
    find patterns in clickstreams of u
    put u to group according to prevailing pattern
for each group g do
    for each user u in group g do
        sort users in group g according to their
        similarity to u
```

Every user ends up in exactly one group according to the most dominant pattern found in his surfing history. There is one more group for users with no dominant navigational pattern. The order of similar users from one group is unique for each user.

Alg. 2 presents the process of recommending links among users.

Algorithm 2 Recommend links for user *u* by similar users.

```
similar = select top K similar users
for each user v in similar do
    for each page p visited by v and not visited by u do
        predict interest of user in page (u, p)
recommend top M pages with highest predicted interest
```

Navigational patterns of users have to be of a certain minimal length (so that each sequence of two following pages does not represent a *path* pattern). After finding similar users to user *u* we select top *K* of them to form a recommendation group. The groups change according to new browsing sessions in which the users can behave differently. This reflects the evolution of user's behavior in time. However, at each time the user belongs to exactly one group according to the most dominant pattern in browsing history.

3.2 Determining interest of users

In order to recommend links to a particular user we need to evaluate the interests of the users in his recommendation group. We can recommend pages which other users liked. To determine user's interest in a particular web page we observe actions he makes on this page. These include *time spent on a web page*, *number of scrolling events* that occur and *number of times he copies text into the clipboard*. We chose these interest indicators because their tracking is platform independent.

Our method is based on the comparison of current user's behavior with the behavior of other users. We compare the values of time and scrolling with values from other people who visited the same page. If the value for the current user is more than X % higher than the average, we consider it as a sign of positive interest in the page. In contrast, when it is more than X % lower than the average we consider it as a sign of negative interest. When the value is around average (± X %) it is a sign of neutral interest. This way we can also consider other interest indicating actions. The exact value of X depends on the calibration for selected domain; in our experiments we used the value of 20 %.

When no behavioral data for a particular web page is available we cannot estimate the user's interest. This is a problem with newly added pages as well as with pages visited for the first time.

We estimate the actual value of user's interest in each page he visits according to Figure 1. We increase this value by 0.1 when the user also copied text into clipboard; otherwise we decrease it by 0.1. Resulting interest is in the interval <0,1> with 0 meaning no interest and 1 meaning total interest in the visited web page.

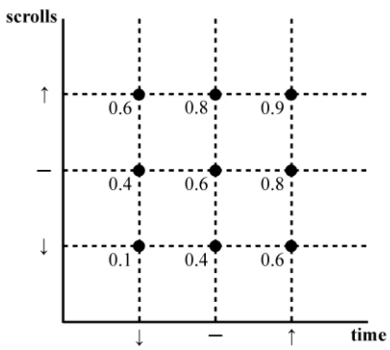

Figure 1. Estimation of user's interest from his actions.

3.3 Social recommendation of links

We recommend web pages by predicting user's interest in yet unseen pages using collaborative filtering method. We compute the predicted value of interest like this [15]:

$$p_{a,i} = r_a + \frac{\sum_{u=1}^{N} (r_{u,i} - r_u) \times S_{a,u}}{\sum_{u=1}^{N} S_{a,u}}$$

where $p_{a,i}$ means prediction of interest of user a in page i, r_a is the average interest of user a in all visited pages, $r_{u,i}$ is the interest of user u in visited page i, $S_{a,u}$ is the value of Pearson correlation coefficient [15] between users a and u determining the similarity of their interests, and N is the number of similar users.

4. EVALUATION AND EXPERIMENTS

To evaluate the proposed method for user's interest estimation we developed software tools which support adaptive navigation by recommending information extracted from potentially interesting web pages to guests of particular web portal. We experimented with the web portal of our faculty (www.fiit.stuba.sk).

We proposed client-server architecture with an adaptive proxy [3] in the middle as shown in Figure 2. Adaptive proxy can be extended to conduct various methods of web personalization on any web portal. We use proxy to put behavior tracking script into the web page. It sends logged behavioral data to the server when the user is active (i.e. when he uses the mouse). The user is aware of this when he agrees to use our proxy server. The data is anonymous – we only know a random ID associated with each user. The delay caused by the proxy server is imperceptible.

One component (*SpyImp*) creates the domain model by analyzing web pages of selected web portal. Another server component (*AdaptiveImp*) is responsible for grouping of users, estimating their interests and making predictions for unseen pages. Then it selects the links to be recommended. The user model consisting of the session vectors and his behavior is being periodically updated.

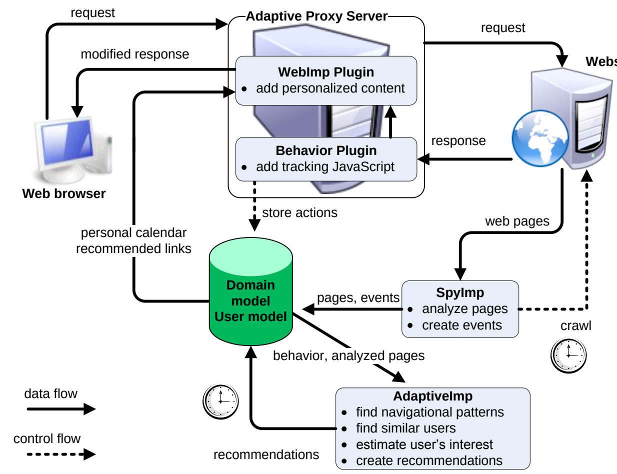

Figure 2. Architecture of proposed link recommender system.

Our plug-in to the adaptive proxy (*WebImp*) modifies the web page by adding special sections with recommended links. One of those sections is personalized calendar. Many web pages on the web portal of our faculty inform about an upcoming event. We automatically extract dates from these pages and create events. Using proposed method we determine user's interest in such page. Then, if the interest is positive, we add the event to user's calendar. This way we also recommend events among users.

We monitor the portal and capture every change in text of a web page (this could be for example a change in time and place of some event). Every changed page is marked as *news* and added to a special news section. We also recommend other potentially interesting links which are neither *events* nor *news*. Figure 3 shows part of a web page enhanced with personalized sections.

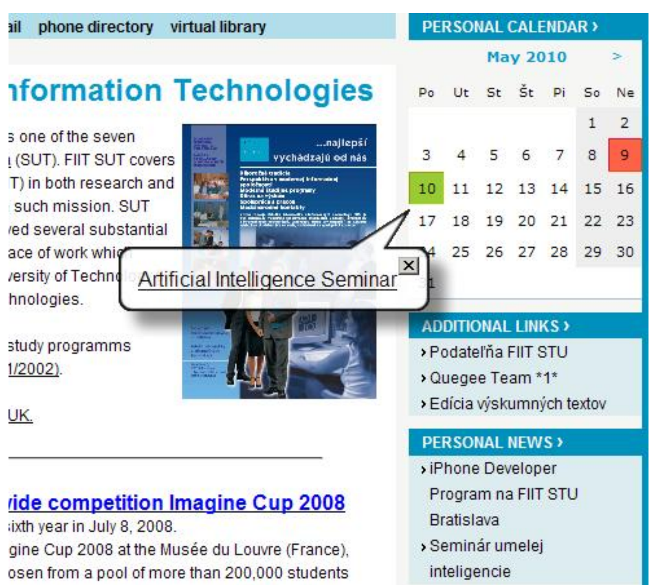

Figure 3. Calendar (shows recommended event on 10/05/2010), additional links and personal news sections.

We provided a series of experiments on our faculty website. Actions of 24 users on a modified website were monitored for 3 weeks. We compared our calculations with their explicit feedback.

Results indicate that time actively spent on a web page is the best interest indicator. Scrolling proved to indicate positive interest as well. However, when the user does not use scrolling, it does not always mean he is not interested in the page. The accuracy of our interest estimation method was 62 %.

The sections with recommended links – especially calendar – were attractive (according to answers from questionnaire) and the users found 55 % of recommended links and events interesting.

Some users were not satisfied with the recommendations. The problem was that they visited the website for the first time. Hence their user model was empty and we could not provide suitable recommendations. This is a common problem with recommender systems and new users. We tried to overcome it by recommending the most interesting events (links) according to behavior of all users. However, this is not always a suitable solution.

5. CONCLUSION AND FUTURE WORK

We have presented a method for adaptive recommendation of interesting links. Our approach is based on collaborative filtering, which has a potential to be used in unusual ways. We presented one of them when considering data about user's actions instead of content of pages. This way we are able to predict user's interest for unvisited pages. Our method achieves solid results and can be further improved in a recommender system which will combine content analysis with behavior, which is our plan for future work.

In this paper we presented a useful application of our method by creating personalized calendar of events on our faculty's website. Using this method we can also personalize other sections of a web page as well. In our opinion recommender systems should bring added value to users by doing further analysis of the domain which is being adapted. On the web they should recommend particular objects (e.g. events) instead of simply listing potentially interesting links to web pages. In order to accomplish this we need to use more text processing algorithms in our recommender systems so that they "understand" the meaning of text on the web.

We ran up against a problem with incorporating the sections with personalized content into a website. In order to do this we need to know the semantics of the website's structure. This is also useful during page analysis and content extraction. We consider the special tags in HTML5 (e.g. *nav*, *footer*) to be insufficient so we came up with a descriptive XML file which pairs HTML tags and their IDs with their predefined meaning (left menu, right menu, etc.). This way our recommender system understands the structure of a website and can alter some sections. We think that there is a need for further development of this format and we see an opportunity for its adoption by other recommender systems.

6. ACKNOWLEDGEMENTS

This work was supported by grants VG1/0508/09, KEGA 028-025STU-4/2010 and the Foundation of Tatrabanka. It is a partial result of the Research & Development Operational Program for the project SMART II, ITMS 25240120029, co-funded by ERDF. We wish to thank members of PeWe group, pewe.fiit.stuba.sk for valuable discussions and their help in experimental evaluation.

7. REFERENCES

[1] Baraglia, R., et al. 2006. A Privacy Preserving Web Recommender System. In *Proc. of the ACM Symposium on Applied Computing*, (Dijon, France). ACM Press, 559-563.

[2] Barla, M. and Bieliková, M. 2009. On Deriving Tagsonomies: Keyword Relations coming from the Crowd. In *LNAI 5796, Proc. of Int. Conf. on Computational Collective Intelligence*, ICCCI 2009, Springer, 309–320.

[3] Barla, M. and Bieliková, M. 2010. Ordinary Web Pages as a Source for Metadata Acquisition for Open Corpus User Modeling. In *IADIS Int. Conf. WWW/Internet.* To appear.

[4] Barla, M., Tvarožek, M. and Bieliková, M. 2009. Rule-based User Characteristics Acquisition from Logs with Semantics for Personalized Web-based Systems. *Computing and Informatics*, Vol. 28, No. 4, 399-427.

[5] Brusilovsky, P. 1996. Methods and Techniques of Adaptive Hypermedia. *User Modeling and User-Adapted Interaction*, Vol. 6, No. 2-3, Springer Netherlands, 87-129.

[6] Canter, D., Rivers, R. and Storrs, G. 1985. Characterizing User Navigation through Complex Data Structures. *Behavior and Information Technology*, Vol. 4, No. 2, 93-102.

[7] Cockburn, A. and McKenzie, B. 2001. What do Web Users Do? An Empirical Analysis of Web Use. *Int. Journal of Human-Computer Studies*, Vol. 54, No. 6, 903-922.

[8] Das, A.S., et al. 2007. Google News Personalization: Scalable Online Collaborative Filtering. In *Proc. of the 16th Int. Conf. on World Wide Web* (Banff, Canada), ACM Press, 271-280.

[9] Gurský, P., et al. 2009. User Preference Web Search – Experiments with a System Connecting Web and User. *Computing and Informatics.* Vol. 28, No. 4, 515-553.

[10] Joachims, T., Freitag, D. and Mitchell, T. 1997. WebWatcher: A Tour Guide for the World Wide Web. In *Proc. of the Int. Conf. on Artificial Intelligence* (Nagoya, Japan), Morgan Kaufmann, 770-777.

[11] Krištofič, A. and Bieliková, M. 2005. Improving Adaptation in Web-Based Educational Hypermedia by means of Knowledge Discovery. In *Proc. of 16th ACM Conf. on Hypertext and Hypermedia*, ACM Press, 184-192.

[12] Milic-Frayling, N., et al. 2004. Smartback: Supporting Users in Back Navigation. In *Proc. of the 13th Int. Conf. on World Wide Web* (NY, USA), ACM Press, 63-71.

[13] Návrat, P., Taraba, T., Bou Ezzeddine, A. and Chudá, D. 2008. Context search enhanced by readability index. *IFIP WCC Series 276*, Springer, 373-382.

[14] Suchal, J. and Návrat, P. 2010. Full Text Search Engine as Scalable k-nearest Neighbor Recommendation System. *IFIP WCC Series 331*, Springer, 165-173.

[15] Sugiyama, K., Hatano, K. and Yoshikawa, M. 2004. Adaptive Web Search based on User Profile Constructed without any Effort from Users. In *Proc. of the 13th Int. Conf. on World Wide Web* (NY, USA), ACM Press, 675-684.

[16] Velayathan, G. and Yamada, S. 2006. Behavior-based web page evaluation. In *Proc. of the 15th Int. Conf. on World Wide Web* (Edinburgh, Scotland), ACM Press, 841-842.