

Layered Evaluation of Adaptive Search

Peter Brusilovsky
University of Pittsburgh
School of Information Science
Pittsburgh, PA, 15260
peterb@pitt.edu

Rosta Farzan
Intelligent Systems Program
University of Pittsburgh
Pittsburgh, PA, 15260
rosta@cs.pitt.edu

Jae-wook Ahn
University of Pittsburgh
School of Information Science
Pittsburgh, PA, 15260
jaa38+@pitt.edu

ABSTRACT

The goal of this paper is to discuss how adaptive search systems should be evaluated. We argue that a state-of-the art evaluation of adaptive search systems should follow a “layered evaluation” approach. To support and explain this argument we describe how layered approach was applied to evaluation of adaptive search component of Knowledge Sea II system that is powered by a social navigation support mechanism.

Categories and Subject Descriptors

H.3.4 [Systems and Software]: Performance evaluation (efficiency and effectiveness)

General Terms

Measurement, Design, Human Factors

Keywords

Social search, adaptive systems, exploratory search systems, layered evaluation

1. INTRODUCTION

The growing need for effective organization and maintenance of increasing Web based educational resources motivated us to construct personalized information access system, Knowledge Sea II (KSII). KSII provides various types of information access methods including two-level visualization (knowledge map plus similarity based visualization), hypertext browsing, recommendation, and social search. Personalization for all these access methods is provided by social navigation support [1], [5]. Social navigation is a relatively well-known personalization approach for browsing-based and recommendation-based information access, however, its use for search personalization is almost unexplored.

The adaptive search component of KSII combines traditional vector search engine with social navigation support allowing every user to benefit from the collective wisdom of the whole community. To stress it we will refer to it as “social search”. Important for the purpose of the paper is that the results of the search are adapted to the user taking into account past interactions

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

SIGIR '06, August 6–11, 2006, Seattle, WA, USA.

Copyright 2006 ACM 1-58113-000-0/00/0006...\$5.00.

of the user herself and the group she belongs to. Main goal of this paper is to discuss how adaptive search systems should be evaluated using our evaluation of KSII search as a case in this discussion. We argue that a state-of-the art evaluation of adaptive search systems should follow a “layered evaluation” approach that is an active focus of research in the area of user-adaptive systems [2]. The core idea behind layered evaluation is that specific sub-components or layers of any user-adaptive system should be understood and evaluated independently. Layered evaluation can overcome shortcomings of the conventional evaluation methodology for adaptive systems which tries to test the adaptation process as a whole and can miss success of failure of critical sub-components. In our approach for the layered evaluation, we divided the adaptation process into two layers: decision making and interface adaptation and then evaluated each of them. In this paper, the nature of our adaptive social search system is presented (section 2) and our layered evaluation framework (section 3) is discussed. The paper concludes in section 4 with summary and a brief discussion of our future direction of our research.

2. SOCIAL SEARCH IN KSII

KnowledgeSea Search c_tutorials

Query: function memory pointer

Stemmed query: function memori pointer | 504 of 2498 documents retrieved | Search time: 0.08 seconds

Visualize: VIBE

Result pages: 1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26

Rank	Source	Title	Score	State
1	D. Marshall	chapter2_12.html	0.4730	
2	C.Faq	s4.html	0.4487	
3	S. Summit	11.2 Freeing Memory	0.4474	
4	S. Summit	11.4 Pointer Safety	0.4454	
5	R. Milles	Memory	0.4382	
6	R. Milles	malloc	0.4148	
7	Univ. of Leicester	Pointers	0.4057	
8	D. Marshall	Pointers	0.3991	
9	D. Marshall	node10.html#ch:pointers	0.3969	
10	R. Milles	c13.html	0.3968	
11	S. Summit	Chapter 11: Memory A ...	0.3955	
12	D. Marshall	section2_12_2.html	0.3947	
13	Landmarks	L21/tslid007.htm	0.3942	
14	C.Faq	Question 1.22	0.3889	
15	D. Marshall	What is a Pointer?	0.3843	
16	D. Marshall	section2_12_1.html	0.3826	
17	D. Marshall	node10.html#fig:float	0.3757	
18	D. Marshall	node10.html#fig:point	0.3738	
19	D. Marshall	Pointer and Function ...	0.3709	
20	D. Marshall	Advanced Pointer Top ...	0.3503	

Contact: Jae-wook Ahn <jaa38 at pitt.edu>

Figure 1 Social search with social navigation cues

Social navigation in KSII powers several information access methods including social search. Social navigation support is offered through visual annotation of the links with icons and color codes. Figure 1 shows an example of search results annotated with social navigation cues. To the right from standard set of information about each document in the list - like rank (7), document source (Univ. of Leicester), title (Pointers), and similarity score (0.4057) - are traffic-based and annotation-based

social navigation cues. The foreground and the background colors of human icon contrast the user and group traffic, which is associated with time spent reading this document [4]. The darker the color is, the higher the traffic. The background color of the annotation represents the annotation density. The foreground icons represent the type of the user annotation and the overall status of the annotation. The “thumbs-up” icon represents positive user annotation and the warm temperature shown by the “thermometer” icon represents positive group annotations for the page. For example, the document “Pointers” shown on Figure 1 is ranked 8th in terms of its similarity score to the user query but is very popular among the community of the users.

3. LAYERED EVALUATION OF SOCIAL SEARCH

The need for the layered evaluation framework was raised from the insight that conventional evaluation methods could not help understand the effectiveness of the critical layers of the adaptation process, which perform different tasks and then contribute to the final results. Current evaluation practices frequently attempt to evaluate adaptation as a whole by comparing the whole adaptive application and non-adaptive application as a baseline. However, even if the result turned out to be better than the baseline, we cannot hastily conclude all of its components perform well. Vice versa, if the adaptive system as a whole failed to beat the baseline, we should not miss the possibility that one of its adaptation layers was actually successful [3]. To address this problem, several authors introduced layered evaluation frameworks. Brusilovsky et al defined user modeling and adaptation evaluation layers in [2]. Weibelzahl introduced a 4-layer approach: reliability and validity of input data, interface, adaptation decision, and interaction evaluation [6].

To evaluate social search in KSII we adopted a 2-layer approach considering separately decision making and adaptation layers in the system. The decision making layer processes the interaction history of all users of the system and decides which pages should be useful from the social prospect for the users of each group and to what extent. The adaptation layer decides how to express the social importance of a specific page to each user. The adaptation layer in the current version of KSII generates icon-based annotations shown on Figure 1 – however it is just one possible way to express social importance of documents estimated by the decision making.

3.1 Decision Making Layer

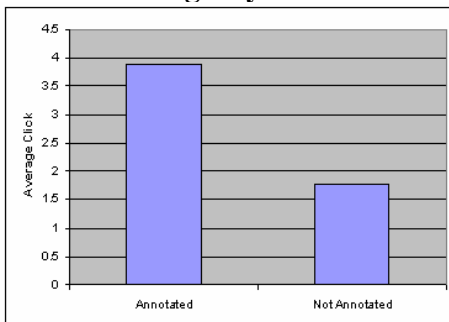


Figure 2 - Average click number over pages with and without annotations

The goal of the decision making layer is to predict how useful each document is to a user of a specific group. KSII uses two independent decision-making layers – traffic based and annotation based. Since the latter is rather straightforward, we focused on evaluating the traffic based one. To argue that the traffic-based prediction works we need to show that documents predicted as useful (those shown with darker blue background by the adaptation layer) are really useful. Our gold standard for rating the importance of pages is by what student find good and important. Therefore, we focused on pages with annotations made by students.

For evaluation, we computed the normalized access rate for pages with and without annotations. As can be seen in Figure 2, “good and important” pages are accessed twice as often. Thus page traffic average is a good indicator of page quality. “Good and important” pages will have a generally darker background, according to our traffic-based social navigation support algorithm.

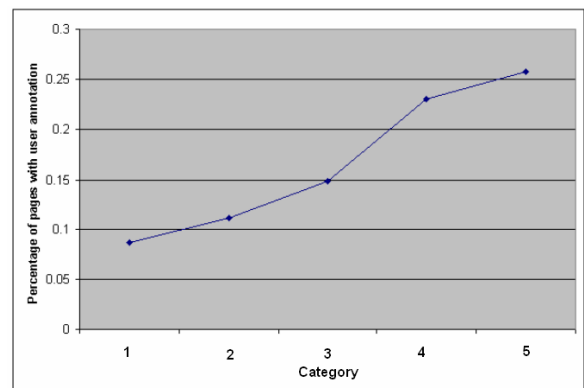


Figure 3 - Percentage of pages with user annotation for different level of usage

To perform enhanced evaluation we categorized accessed documents based on the time spent on each page into five categories. The following table shows the details of our classification.

Category	Average Time Spent	Darkness of Background	Level of Recommendation
1	< 65 sec	No background	Not recommended
2	< 97 sec	Light blue	Slightly recommended
3	< 152 sec	Blue	Recommended
4	< 217 sec	Dark blue	Considerably Recommended
5	> 282 sec	Very dark blue	Highly Recommended

For each category we computed percentage of pages that are annotated by the students. To exclude the dependency of annotation and visit, we excluded annotations made by the users of the target semesters and included annotations made by users of past and future semesters. As shown in figure 3, the pages with darker background (higher number of usage) have higher percentage of annotation. The data shows that important pages are being predicted as useful by our social navigation adaptation

which means the important pages are augmented with darker background.

3.2 Evaluation of social search with social navigation cues

Once we established the positive correlation of quality and social navigation, it is important to evaluate the effect of social navigation cues. The goal of the cues is to attract user attention to socially important documents and to encourage them to examine them. In our context, we need to evaluate how much the social navigation cues affect students' decision in choosing a link from search result. Moreover, since KSII social search separates the visualization of query relevance (shown by document position in the search list) from visualization of social importance (shown by intensity of background color in social navigation cues), we were interested to compare the "attractiveness" of social navigation cues with the "attractiveness" of top search rank.

To evaluate this layer, we decided to compare effective and random relative access rate for links with low rank (on the top of the list) and links with traffic-based cues. The *random relative access rate* tells which fraction of clicks will be made on links with a specific property if the user selects links in the search results list randomly. Basically, it shows how often the links with this property appear in the search results list. The *effective relative access rate* reports the real fraction of links with the target quality among all accessed links. If the effective relative access rate is higher than random, it means that the links with this quality encourage user to access them.

The first question to answer is "do students prefer links with better rank" (we consider top three documents in the search results list with rank 1 to 3 as top ranked). Since every result page shows 20 links, the random relative access rate for top three ranked documents is $3/20 = 0.15$. Effectively, students accessed 53 documents from different search results lists and out of these 53, 16 were among the top 3 documents. Therefore the effective relative access rate is $16/53=0.3$, which is twice the random (0.15). This is evidence that the students do take the document rank into account preferring links on the top of the list.

The second question to answer is "do students prefer links with traffic-based social navigation cues". Answering this question we attempted to evaluate separately links with any visible past traffic (number of past clicks >1) and links with higher traffic (number of past clicks >2). The reason is that the links with two past click were annotated with a very light blue color, which, we afraid, some users might ignore. The links with 3 and more past clicks were annotated with reasonably dark blue color and were hard to ignore.

Computing random relative access rate for links with group traffic was a complicated procedure. For each of the 53 cases of link access we had to re-create the group traffic accumulated at the time of access to understand how many links with social cues the user saw when making the selection. For each case we calculated the random relative access rate by dividing the number of links with target level of traffic to the total number of links. Then we averaged the probabilities over all 53 cases and found that for pages with visible traffic the random relative access rate is equal to 0.08. Out of 53 cases, students choose 17 documents from the

visible traffic category. Therefore the effective relative access rate for links with visible traffic is $17/53=0.32$, which is four times higher than random access (0.08). Similar ratio (0.05 to 0.19) was obtained for links with high traffic. This result shows that students do prefer links with visible group traffic. Moreover, the ratio of their effective access rate to random is twice as higher for pages with visible traffic than for pages in the top of the results list. It provides evidence that pages marked by visible group traffic do attract student attention. Moreover, a presence of a "group traffic" give a page a higher chance to be visited than even page location among the top three results returned by the search engine.

4. CONCLUSIONS

In this study, we demonstrated how a 2-layered evaluation framework could be used for evaluating an adaptive search interface. We divided the whole evaluation process into decision making and adaptation layer in order to better understand the effectiveness of the process. We were able to show correlation between predicted and effective social utility of a page (i.e., pages predicted as important for the group by the decision making component were really important). We also provided evidence that the specific interface adaptation approach used in KSII to attract user attention to socially important pages does influence user behavior in the expected direction. The proposed framework will be able to evolve in terms of granularity by adopting more layers such as user to system interaction evaluation and validity of input data evaluation. As our future work we are planning to evaluate other kinds of adaptive information access methods including information visualization with the same layered framework.

5. REFERENCES

- [1] Brusilovsky, P., Chavan, G., and Farzan, R (2004). Social adaptive navigation support for open corpus electronic textbooks. In Proc. of Adaptive Hypermedia and Adaptive Web-Based Systems.
- [2] Brusilovsky, P., Karagiannidis, C., and Sampson, D. Layered evaluation of adaptive learning systems. *Int. J. Cont. Engineering Education and Lifelong Learning*, 14, 4, 2004, 402-420.
- [3] Brusilovsky, P., Karagiannidis, C., and Sampson, D. (2001). The Benefits of Layered Evaluation of Adaptive Applications and Services. In *Proceedings of Workshop on Empirical Evaluation of Adaptive Systems*.
- [4] Farzan, R. and Brusilovsky, P. (2005) Social navigation support in E-Learning: What are real footprints? In Proceedings of IJCAI'05 Workshop on Intelligent Techniques for Web Personalization.
- [5] Farzan, R. and Brusilovsky, P. (2005) Social navigation support through annotation-based group modeling. In Proc of User Modeling.
- [6] Weibelzahl, S., and Weber, G (2001). Advantages, Opportunities, and Limits of Empirical Evaluations: Evaluating Adaptive Systems. *Kunstliche Intelligenz*.
- [7] Lawrence, S., and Giles, C.L. Context and Page Analysis for Improved Web Search (1998). *IEEE Internet Computing*.