Searching the Web with a Little Help from your Friends

Taly Sharon, Henry Lieberman, Ted Selker

MIT Media Lab
20 Ames St
Cambridge, MA 02139 USA
{taly, lieber, selker}@media.mit.edu

ABSTRACT

When browsing the Web, users have limited knowledge about the quality of the information they view. Implementing a system to address this issue usually requires active information gathering and the changing of user work habits. We propose a system for augmenting Web pages with annotations. It is based on access statistics and requires no input from the users. Preliminary experiments indicate that it has good potential to support effective ranking for quality judgment by users.

Keywords

Recommender system, collaborative filtering, user interface

INTRODUCTION

When browsing the Web, what quality information does the user have about the information underlying the presented links? The only existing hint behind links is that they change their color when visited. However, a lot of relevant knowledge that is possessed by the user's colleagues is not exploited. The aim of this paper is to present a system for enhancing links in Web pages, based on the experience of the community, i.e., members in the same organization.

MOTIVATION

Assume a user is searching for information. She can use a search engine, but she will get the broad viewpoint of Internet users and content providers. She can use methods such as collaborative filtering (also called Value-ranking) [6] to take advantage of other people's experience. However, that view is not focused enough for her specific context: her profession, organization, location, etc.

An alternative is to find a peer in the organization and ask her for recommendations. However, it takes time and energy of both sides. Moreover, it might not be clear to her who is the right person to ask, that person may not be available, or might not recall the information immediately.

For this reason, collaborative filtering that is based on peers and colleagues in the same organization has a better potential to be useful. Furthermore, discovering the opinion of an expert in the same organization has a significant added value. Thus, the experts should be identified and their knowledge should be put to good use.

We want to gather up-to-date, quality information from users in the organization, while identifying their fields of expertise. Asking users to fill out forms or annotate pages is impractical. Firstly, users hate forms; they cannot be trusted to fill them in correctly [5]. Secondly, annotating is a tedious work; people will not cooperate. Thirdly, even a simple task such as uploading Bookmarks will require updating them, a non-easy task

Considering this, our goal is to provide a Zero-input interface - an interface that does not require any explicit user input [4]. The data should be gathered as a side effect of the regular work of people in the organization - not by any explicit action. Similarly, the consumers of this information need not change any of their working habits in order to use the system.

Furthermore, the system should be aware of the context of the user task [5]. When the user task is browsing – it should enhance links. When the task is searching – it should provide additional links. The enhancement should be made on the same page that the user views, and not in a separate interface. Furthermore, the user should not be provided with loads of information. Hence, providing small symbolic information while expanding it based on mouse-moves seems the best option.

IMPLEMENTATION

We propose a Context-aware Annotation Proxy-based System (CAPS), a system for "on-page" link annotation. CAPS acquires its information via a Zero-input Interface (explained below) using a proxy. Then, it annotates the pages according to the user task. We demonstrate the system by suggesting the following scenario: a user is trying to find information about a certain subject. The annotations and recommendations provided by the system help her find high quality information and so save valuable time. CAPS is a mediator, written in Java under WBI [1], behaving like a proxy with some changes:

1. It logs user requests, constructing a repository of access statistics and page metadata.
2. It builds user profiles based on keywords appearing in their homepages.
3. It ranks pages based on these access statistics and a matching with the expert profiles (based on points 1 and 2 above).

4. It uses this information to augment browsing via a unique user interface (described below).

User interface

A user sets up CAPS by defining it as a proxy in the browser. If the user wishes to be considered an expert in the system, she records this by providing CAPS with her homepage URL. This is actually the only explicit interaction required of the user when using CAPS.

When the user browses the Web, CAPS slightly modifies the pages to reveal the popularity of the links. Each link recognized by CAPS is annotated with a small image next to it. The image reveals the score of the link and is presented in one of three levels: "known" (yellow), "popular" (yellow-red) and "hit" (red). Hovering above these images will extend the information presented in a ToolTip style. For an example of the CAPS user interface see screen shot in figure 1.

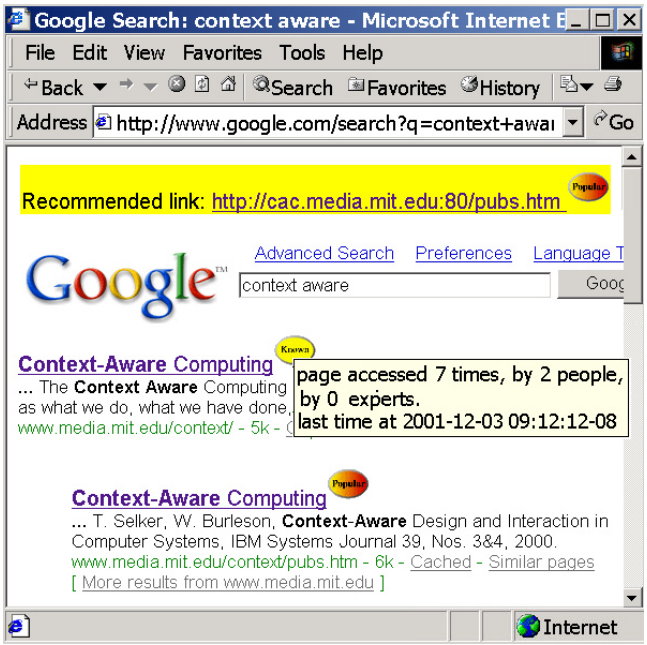

Figure 1: CAPS user interface

Whenever accessing one of the popular search engines, CAPS adds a highlighted recommended link at the top of page and ranks it. The link itself changes in every access, to reveal new information to the user.

EVALUATION

We conducted an experiment to evaluate CAPS ranking. A group of users reported their browsing habits using their history files. We used this report to create a combined list of sites, and ranked them using CAPS's algorithm. Then we asked a larger group of subjects from the same organization to comment on the ranking. We found that 3% were ranked too high, while 7% of the sites were ranked too low. Thus, CAPS achieved a ranking accuracy of 90%.

We also asked a group of people to use CAPS and refer to its usability. All the subjects liked the non-intrusive user interface and the links annotations. Some wished it had a more explicit search interface to the CAPS repository.

RELATED WORKS

Previous works that used history and proxy data to build a document repository or to provide page ranking required the users to actively provide input or give feedback [6].

An interesting system, based on proxy architecture [2], aimed to support cooperative browsing. It provided awareness to other group members logged into the proxy by listing the links they visited. The system highlighted links to pages that were currently cached in the proxy while marking fetching errors. The system enabled the group to communicate via Chat. While providing information about the activity of the group, assessing links quality required the asking of group members using the Chat, and assuming they are available online. Thus, this architecture did not facilitate automatic quality assessment of the links.

Surflen [3] implemented a Zero-input interface to recommend similar pages based on clusterization of users and their browsing history. It used a separate monitor to present the recommendations, while CAPS provides both links ranking and recommendations on the same page. In addition, CAPS's algorithm is different than the one used in Surflen, and also from classic collaborative filtering. CAPS makes use of the users' history to provide ranking of the sites; its recommendations are based on a search query and their results ranking, rather than on document and users similarity. Furthermore, the use of a statistical combination of numbers of access times, users, and experts (based on their homepage), makes its ranking algorithm unique.

CONCLUSIONS AND FUTURE DIRECTIONS

In conclusion, browsing history, combined with expertise information, has a potential to provide a Zero-input interface for Web pages ranking and recommendation.

Future work should concentrate on testing the system more thoroughly and on tuning the algorithm as well as exploring ways to improve the presentation.

REFERENCES

1. Barrett R. & Maglio P.P., Intermediaries: An approach to manipulating information streams, in *IBM Systems Journal* 398, 629-641, 1999.
2. Cabri G, Leonardi L., Zambonelli F, Supporting Cooperative WWW Browsing a Proxy-based Approach, *7th Euromicro Workshop on Parralel and Distributed Processing,* 138-145, February 1999.
3. Fu X., Budzik J, Hammond K.J., Mining Navigation History for Recommendation, *IUI*: 106-112, 2000.
4. Lieberman H., Fry C. and Weitzman L, Exploring the Web with Reconnaissance Agents, *Communication of the ACM Journal* 44(8):69-75, 2001.
5. Lieberman H. and Selker T., Out Of Context: Computer Systems That Adapt To, and Learn From, Context, in *IBM Systems Journal* 39, Nos. 3&4, 617-632, 2000.

6. Paepcke A, et. al., Beyond Document Similarity: Understanding Value-Based Search and Browsing Technologies, *in SIGMOD Records*, 29(1), 2000.