

Massive Implicit Feedback: Organizing Search Logs into Topic Maps for Collaborative Surfing

Xuanhui Wang and ChengXiang Zhai
Department of Computer Science
University of Illinois at Urbana-Champaign
{xwang20, czhai}@illinois.edu

1. INTRODUCTION

Current search engines heavily emphasize on direct querying which tends to work well only for simple information needs such as navigational queries. However, direct querying may not support complex information needs such as exploratory search well [4, 11] since users' interactions are mainly limited to submitting a query, viewing results, and reformulating queries [1]. As a complementary way of information seeking with querying, browsing can be very useful for exploratory search or information foraging [5]. Unfortunately, with the current search engines, browsing is mostly limited to following hyperlinks or navigating through structures consisting of a fixed set of categories or other meta-data available [4, 2].

We have been developing a new collaborative surfing system to enable users to go beyond hyperlinks to browse flexibly for *ad hoc* information needs. Our main idea is to view search logs as information footprints left by users in navigating in the information space and organize these footprints into a multi-resolution topic map. The map makes it possible for users to navigate flexibly in the information space by following the footprints left by other users. As new users use the map for navigation, they leave more footprints, which can then be used to enrich and refine the map dynamically and continuously for the benefit of future users. Thus, by turning search logs into a topic map, we can establish a sustainable infrastructure to facilitate users to surf the information space in a collaborative manner. Preliminary experiment results show that the topic map is effective in helping users to satisfy exploratory information needs [8]. In the following, we describe our system in more detail and discuss its potential impact on understanding users for improving information seeking.

2. SYSTEM DESCRIPTION

Figure 1 shows the interface of our system which is implemented as a meta-search engine interacting with Google. The interface has three panes: (1) The top pane is a query-

ing box, where a user can submit a keyword query. (2) The left pane shows a portion of a multi-resolution topic map built based on search logs, where a user can click on a node to navigate into a topic region. (3) The right pane displays information corresponding to a topic region, including the clickthroughs made by previous users when they visit the topic region and the documents covered by the topic region.

These three panes allow a user to navigate in the information space in large, medium, and small steps, respectively. With the query box on the top, a user can make a long distance navigation into any topic region (i.e., "large steps"); with the topic map on the left pane, the user can navigate into related topic regions (i.e., "medium steps"); and with the display of a topic region in the right pane, the user can navigate by following hyperlinks (i.e., "small steps"). A user can take any of these three navigation actions at any time. Thus our system implements a unified information seeking model where both querying and browsing are viewed as ways to navigate in the information space.

Inside the system, when a user submits a query, the system would display the most relevant part of the topic map on the left pane and show the search results from Google for the query. When the user navigates on the map to click on a node (corresponding to a topic region), the system would automatically update the right pane to show corresponding search results using a query constructed based on the node selected by the user. In general, the right pane is always synchronized with the left pane to show the documents corresponding to the current node on the map.

The topic map promotes browsing and can naturally support exploratory search. For example, a user who wants to arrange a house can start with a query "table," zoom into "dinning table," zoom out to "dinning," move horizontally to "kitchen," and further move to "appliance." From "table", this user can also horizontally move to "chair," to "desk," or to "tablecloth." Another example is "wedding." From "wedding," we can zoom into different aspects of wedding such as "wedding dress," "wedding vows," etc. We can also horizontally move to "vacation," "honeymoon," or "hotels." All these browsing traces can be leveraged to infer users' underlying information needs and better serve users with complex exploratory information needs. The browsing logs can be leveraged to improve the map and further help future users who have similar information needs.

A main technical challenge in developing this system is to construct topic maps. Currently, the nodes in topic maps are valid queries in search logs. All queries with the same number of keywords belong to the same level. The children

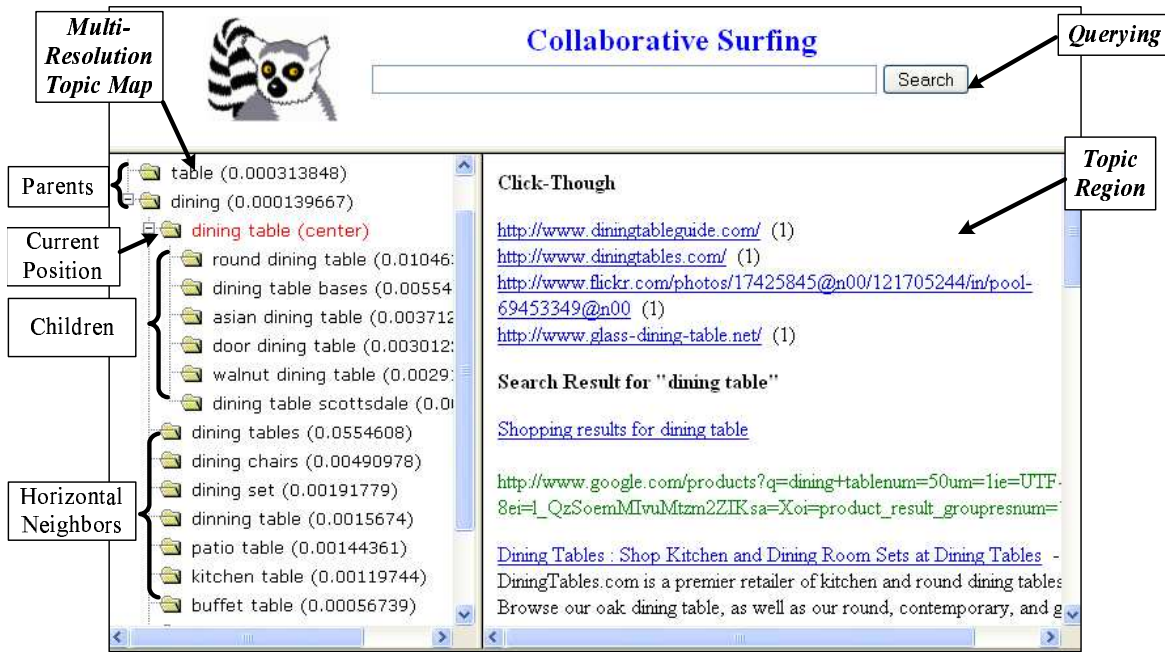


Figure 1: Interface snapshot of the demo system.

of a map node is obtained by adding a keyword into the current query and the neighbors of the query is by substituting a keyword in the current query. All these surrounding nodes are ranked accordingly. Specifically, we rely on the term co-occurrence in search logs to construct such a map and all the technique details can be found in [10].

3. MASSIVE IMPLICIT FEEDBACK

From the viewpoint of understanding users and exploiting user information to provide better search support, our system implements a strategy of massive implicit feedback [3, 7, 6, 9], where query logs and browsing logs of all users would be captured and leveraged to provide better support for future users in both querying and browsing. Indeed, the implicit feedback information collectable by the system includes not only the queries and clickthroughs available in a current search engine but also the browsing traces left by users in using the map. The system treats all these different kinds of user information uniformly as “information footprints” left by users and organizes them into a topic map to deliver benefits to future users. At the same time, new users would leave new footprints to allow the system to grow continuously over time to improve its support for browsing and querying. Thus, the system enables collaborative surfing where users help each other through sustained massive implicit feedback.

We hope our demo can stimulate discussions about many interesting questions related to the workshop: (1) How should we evaluate topic maps? (2) How should we evaluate such an interactive system? (3) How can we formally model a user based on both query logs and browsing logs? (4) How can we leverage maps to clarify user interests?

4. REFERENCES

[1] N. J. Belkin, S. T. Dumais, J. Scholtz, and R. Wilkinson. Evaluating interactive information

retrieval systems: opportunities and challenges. In *CHI Extended Abstracts*, pages 1594–1595, 2004.

[2] M. A. Hearst. Clustering versus faceted categories for information exploration. *Commun. ACM*, 49(4):59–61, 2006.

[3] D. Kelly and N. J. Belkin. Display time as implicit feedback: understanding task effects. In *Proceedings of ACM SIGIR 2004*, pages 377–384, 2004.

[4] G. Marchionini. Exploratory search: from finding to understanding. *Commun. ACM*, 49(4):41–46, 2006.

[5] P. L. T. Pirolli. *Information Foraging Theory: Adaptive Interaction with Information*. Oxford University Press, June 2004.

[6] X. Shen, B. Tan, and C. Zhai. Implicit user modeling for personalized search. In *Proceedings of CIKM 2005*, pages 824–831, 2005.

[7] J. Teevan, S. T. Dumais, and E. Horvitz. Personalizing search via automated analysis of interests and activities. In *Proceedings of ACM SIGIR 2005*, pages 449–456, 2005.

[8] X. Wang, B. Tan, A. Shakeri, and C. Zhai. Search logs as information footprints: Supporting guided navigation for exploratory search. Technical Report UIUCDCS-R-2008-3001, University of Illinois, 2008. <https://www.ideals.uiuc.edu/bitstream/handle/2142/10971/UIUCDCS-R-2008-3001.pdf>.

[9] X. Wang and C. Zhai. Learn from web search logs to organize search results. In *Proceedings of SIGIR 2007*, pages 87–94, 2007.

[10] X. Wang and C. Zhai. Mining term association patterns from search logs for effective query reformulation. In *CIKM*, pages 479–488, 2008.

[11] R. W. White and R. A. Roth. *Exploratory Search: Beyond the Query-Response Paradigm*. Morgan and Claypool, 2009.