Internet Search Using Adaptive Visualization

Dmitri Roussinov
Department of MIS, University of Arizona, McClelland Hall 430
Tucson, AZ 85721
dmitri@bpa.arizona.edu

ABSTRACT
Automatically created maps of concepts improve navigation in large collections of text documents. My research in progress on leveraging navigation by interactively providing the ability to modify the maps themselves has led me to believe that this functionality increases responsiveness to the user and makes searching more effective. I explored both what adaptive features users perceive to be most helpful and the overall effect of adaptation on achieving information seeking goals.

Keywords
Intelligent searching, interactive data exploration, information representation, World Wide Web, search engines, information retrieval, clustering, self-organizing maps.

INTRODUCTION
Summarization and visualization tools have been shown to be helpful in navigating large volumes of data [7]. Visual representation facilitates query reformulation and provides better feedback to a retrieval system [5]. However, automatically created concepts often do not meet users' expectations [2]. In the case of noisy input, like the Internet, automatic summarization sometimes produces results of low quality.

My research seeks to extend the visualization approach to information search described by previous HCI researches [1,2,3]. In this paper, I report my research in progress aimed at improving information visualization by adding an adaptation component. By *adaptation* I mean a user' s ability to modify the way in which the system visualizes a collection of documents.

My approach is based on maps of concepts obtained by document clustering, and automatic assignment of labels to the recognized clusters of documents. Besides having the ability to navigate the semantic categories offered by various commercial and research systems, my system (ASOM) allows users to change the categories themselves. The visualizing system thus adapts to the particular user and the particular search task. Through interaction, the system tries to offer categories that are more meaningful to the user, who thereby is more likely to use the concept map to reformulate the task or navigate the returned documents.

ADAPTIVE VISUALIZATION FEATURES
My prototype system (ASOM) creates a hierarchy of concepts from documents returned by a WWW search engine. It does it by building Kohonen's self-organizing map (SOM) [6]. It may rebuild the map in real-time if asked by the user.

My poster presents a diagram of interaction between the user and ASOM. With ASOM, the search process consists of two general steps: 1) refining the concept map without changing the set of visualized documents; and 2) using the concept map to navigate the retrieved documents or reformulate the query. Step 1 is not present in currently available search systems.

The following adaptive features can be requested through the global menu called "Modify Map" or a local (region-sensitive) menu:

"Remove this term" It removes the specified term from the vector space representation, so it does not show up on the map.

"Use more specific terms." This command rebuilds the whole map and creates concepts using more specific terms.

"Create fewer concepts". The system rebuilds the map using a smaller grid. Typically this leads to fewer regions created on the map.

Each of the above features has the corresponding reverse. A demonstration of ASOM is available on the Web at: http://ai.bpa.arizona.edu/resume/dmitri/query.html

EVALUATION
I performed a pilot study based on a protocol analysis of one search session and analysis of records of 50 search sessions in order to see if the proposed approach is viable.

In my protocol analysis study, the subject was a faculty member interested in grants available to MIS researchers. She first tried the Alta-Vista search engine and entered "MIS grants" using simple query syntax. She received a page claiming that 36952 documents were relevant. My user was not patient enough to go through them. She checked the first 20, judged most of them irrelevant to her search objective and gave up using the search engine directly. Then she resorted to ASOM.

She entered the same query in a simple HTML form: "MIS grants." ASOM built a map for the first 200 documents in 40 seconds (figure 1). My user noted that the terms *mis, management, data* looked relevant to her topic of interest. But she also said that they did not help to focus in on relevant documents, since in this particular context they did not convey any additional meaning. Since the query was about MIS grants, terms like *MIS* and *grants* looked trivial and not very helpful. All documents were likely to be about MIS and grants anyway. She asked for another map to be built without those terms.

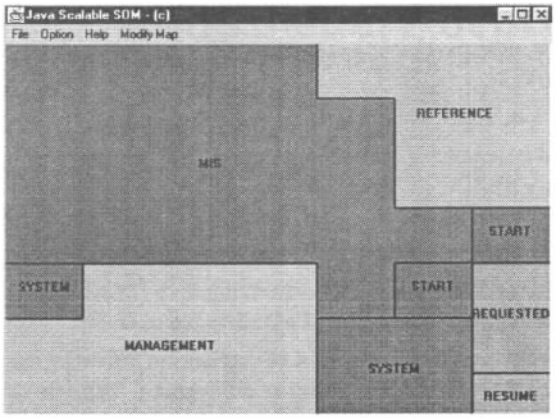

Figure 1: The initial map of categories

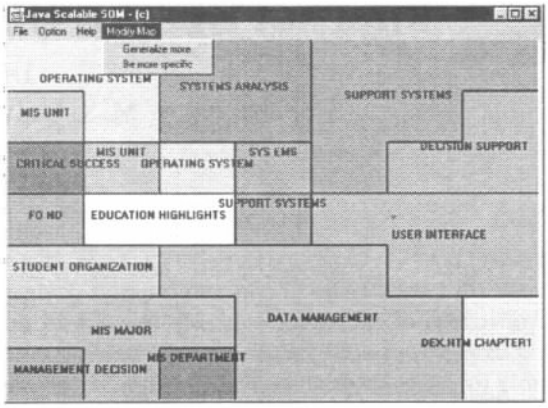

Figure 2: The final map of categories.

The user judged the second map to be more meaningful, but considered the categories to be too general and asked the system to use more specific terms. The resulting map is shown in figure 2. This new map had 12 categories. The user agreed that at this point she could grasp a high-level picture of what concepts were present in the collection of pages presented by search engine. She requested documents belonging to the categories "Operating Systems", "Decision Support", and "User Interfaces".

In the 50 recorded sessions, the majority of users were web surfers who browsed the web site of my laboratory and encountered the Interactive web search demo. I analyzed the traces of 50 search sessions recorded in the period from April 1 to September 10, 1998.

I found that, based on the category maps, users requested not only the top 10-30 documents as ranked by the search engine, but all 200 top documents presented by ASOM with almost uniform probability. This suggests that category maps promote access to documents even beyond the top 10-30.

In 33 out of 50 sessions analyzed, users used at least one of the adaptive features described above. For 113 queries performed in all sessions, users requested adaptive features 118 times, 2.36 times per session in average. This suggests that adaptive features were indeed used. As indirect evidence of search *success rate* I would like to suggest that the proportion of sessions in which users requested documents serve as a proxy. From previous research, users who dislike search result summaries tend to abandon a query rather than investigate the full-text results. From my log files I found that users requested documents in 38% of the sessions.

In 18% of all sessions, users performed a query refinement after looking at the map. 88% of all query refinements were performed after adaptive features had been used. This indirectly indicates that adaptive features help to make refinement decisions.

CONCLUSION AND FUTURE RESEARCH

I have found that users employ and have success with the adaptive features. In order to validate the proposed approach completely, controlled experiments are necessary and underway. First, I would like to test the hypothesis that the adaptive component helps to make category maps for meaningful. This can be done by evaluating maps by experts before/after refinement. Protocol analysis will elucidate what strategies human subjects follow while refining the maps.

The second experiment will test the extent to which the adaptive component helps users to achieve their information seeking goals. The subjects will perform very general search tasks (for example, "Find a camp cite in Yellowstone that is not too cold for November," "Find out how many people live in the capital of Saudi Arabia.") Search efficiency will be measured as time spent on the search task. In addition, the quality of answers (search effectiveness) can be evaluated by expert subjects.

The proposed technique can be applied to other text visualization tasks such as for visualizing web sites, collections of email messages, digital libraries, or electronic brainstorming sessions. The prototype illustrating adaptive text visualization may therefore serve as a good straw man for data exploration in areas overloaded with information.

REFERENCES

1. Card, S.K., Robertson, G.G., & York, W. The WebBook and the Web Forager: An Information Workspace for the World-Wide Web. *Proceedings of CHI '96*, 111-119.
2. Chen, H., Schuffels, C., & Orwig, R. Internet Categorization and Search: A Self-Organizing Approach. *Journal of Visual Communication and Image Representation* 7, 1, 88-102.
3. Cutting, D.R., Karger, D.R., Pedersen, J.O., & Tukey, J.W. Scatter/gather: A cluster-based approach to browsing large document collections. *Proceedings of SIGIR' 92*, 318-329.
4. Hearst, M.A., & Pedersen, J.O. Reexamining the Cluster Hypothesis: Scatter / Gather on Retrieval Results. *Proceedings of SIGIR' 96*, 76-84.
5. Hearst, M.A. Interfaces for Searching the Web. *Scientific American*, March 1997, 68-72.
6. Kohonen T. Self-Organizing Maps. Springer-Verlag, Berlin, Heidelberg, 1995.
7. Pirolli P., Schank P., Hearst M. A., Diehl C. Scatter/Gather Browsing Communicates the Topics Structure of a Very Large Text Collection, *Proceedings of CHI '96*.