

S³: Storable, Shareable Search

Meredith Ringel Morris, Eric Horvitz

Microsoft Research
Redmond, WA, USA
{merrie, horvitz}@microsoft.com

Abstract. We present S³, a system that implicitly captures the process and products of Web *investigations* (exploratory searches involving multiple queries). This automatically-created, persistent representation of an investigation enables future review and continuation of suspended search activities. This persistent representation can reduce unnecessary re-execution of queries and enable users to quickly regain the context of a resumed activity. Stored investigations can also be shared with, and augmented by, collaborators. Furthermore, a stored investigation can act as a *standing query*, proactively updating itself when a user revisits it.

Keywords: Web search, exploratory search, investigation, persistent search

1 Introduction

Web searches are often exploratory or informational [3] in nature. Such *investigations* can involve issuing multiple queries to a search engine and reviewing a large number of resulting web pages. Browsers and search engines, however, typically model search as a transient activity, treating each query independently from prior queries even though, from the user's point of view, multiple queries are frequently part of a single investigation.

This transient model of Web search can result in extra work for users. For example, research on interruptions [4, 5] has found that users frequently switch tasks, and often experience long delays before task resumption. Intervening tasks may also result in closing or changing the state of the user's Web browser. A user returning to a Web investigation after a delay must remember the state of his task, such as what queries he has already issued and what useful sources of information he has already discovered.

The challenges of recalling the state of a resumed task can result in unnecessary duplication of effort. For example, surveys and Web log analyses [1, 8] reveal that users frequently re-enter previously issued queries in order to re-find information. Another study found that over half of Web page visits are re-visits [7].

To address the limitations of the transient nature of current Web search, we introduce the S³ system (Figures 1 and 2). S³ implicitly stores information about the *process* (queries issued) and *products* (useful pages found) of Web investigations. This persistent representation can facilitate resumption of an interrupted or suspended investigation, proactively update itself, and enable collaboration.

Prior approaches to managing complex search tasks include bookmarking, histories, or systems that allow users to flag pages or parts of pages for inclusion in a workspace [2, 6]. Our approach differs from this prior work, since we implicitly store several types of metadata associated with an investigation (queries issued, pages visited, contributor identity, and annotations) and use this stored data to facilitate task resumption, proactive information fetching, and collaboration.

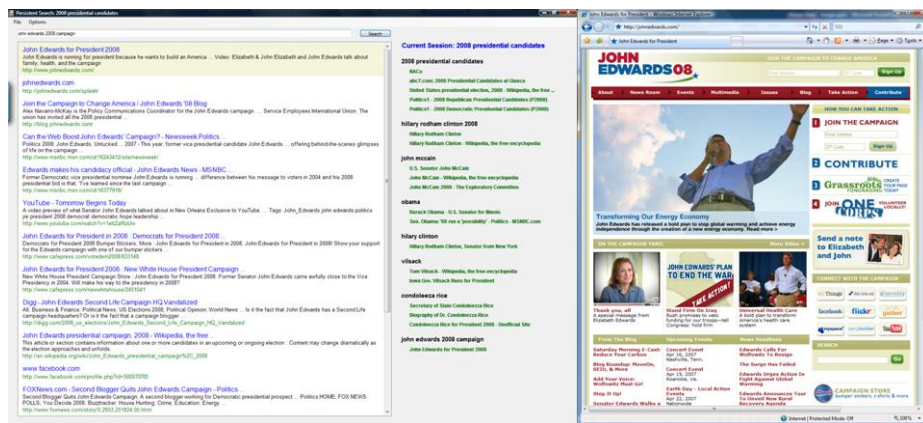


Fig. 1. Conducting an investigation with S^3 . The search box and current search results are shown in the leftmost portion of the interface. The central pane provides an overview of the queries issued and sites visited so far during the investigation. The rightmost portion displays the currently selected result in a browser.

2 The S^3 System

When a user opens S^3 , he can either begin a new investigation by entering a query in the search box (Figure 1), or he can load a saved investigation via the file menu (Figure 2). When a user begins a new investigation, the first query issued is used to provide a default name for the investigation. The user can issue queries via the search box; searches are sent to a Web search engine and the results are shown below the search box (our implementation uses Windows Live Search, but S^3 's architecture allows for the substitution of other search sources). Clicking on any result displays it in a browser window.

For each investigation, S^3 automatically records all queries entered, results retrieved, and subsequent webpages visited. Additionally, users can associate comments with a search result by right-clicking and choosing the “Comment” option from a context menu.

While searching, a user can see a summary of his current investigation history, including queries issued and pages visited, which is displayed next to the current list of search results (Figure 1). At any time, a user can click on this history panel to see a more detailed representation of his investigation (Figure 2) – this representation is also what a user sees if he opens a previously stored investigation.

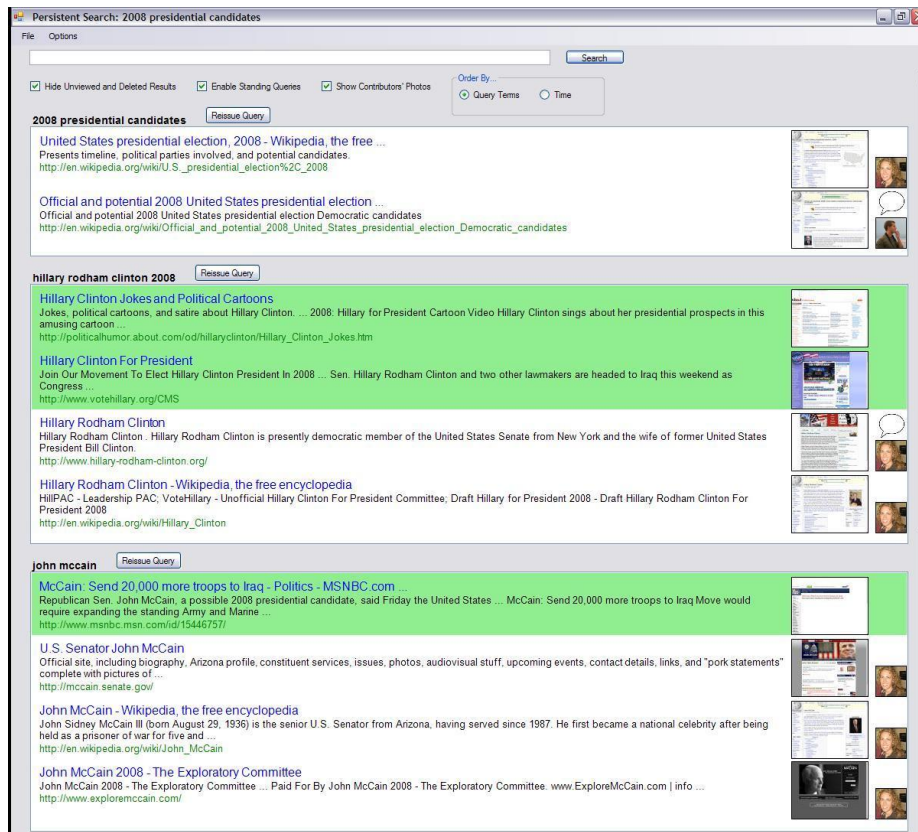


Fig. 2. Viewing a stored investigation. The title, url, and thumbnail for each page visited during the investigation is presented beneath the query terms that lead to its discovery. Items highlighted in green represent newly available content that has been proactively fetched via a standing query. Clicking any item opens the corresponding page in the browser window. The user who contributed each page to the investigation is depicted via a thumbnail, and the presence of comments is indicated by a speech bubble icon; hovering over the icon reveals the full comment.

The detailed view of the investigation shows each query issued during the investigation, in chronological order. Below each query, S^3 shows the title, URL, and thumbnail images of the “useful” pages found as a consequence of that query. The current implementation of S^3 counts any page the user viewed in the browser window as “useful” (and provides the ability to delete a page from the “useful” list via a right-click context menu); however, a variety of heuristics, such as the length of time spent viewing a page or whether links within the page were clicked, could also be used.

A “speech bubble” icon is shown next to pages that have comments associated with them; hovering over this icon with the mouse reveals the full text of the comment. Asynchronous collaboration among multiple users is enabled by S^3 's document model – a stored investigation is saved as an XML file, which can be sent to others (*i.e.*, by email or other file-transfer protocols) and loaded into their own S^3

clients. If more than one user has contributed to the investigation, the users' photos are shown next to the pages that they discovered.

If the user selects the "enable standing queries" menu option, then whenever he loads an investigation S^3 will invisibly re-issue each of the investigation's queries, and check whether each has any highly-ranked results (rank ten or higher) that were not among the top ten results when the user initially executed the query. If such new results exist, then S^3 displays them along with the previously-visited "useful" pages associated with each query; the proactively-fetched results are visually distinguished by green highlighting.

Clicking on any of the pages listed in the detailed view opens that page in a browser, and clicking any of the query terms listed re-executes that query and returns the user to the "searching" view (Figure 1). The search box is also available in the detailed view, enabling the user to add to a stored investigation – entering a query in the search box also returns the user to the "searching" view.

3 Conclusion

The S^3 system can assist users in performing multi-query Web investigations by automatically saving the process and products of a series of related searches. This persistent representation of search aims to help users resume an investigation after an interruption and reduce the need to re-issue queries or re-find sites. The abilities to share investigations with others and to create standing queries are further benefits of our persistent search model. We are currently pursuing studies to evaluate the utility of our persistent search representations in assisting users with multi-session and multi-user searches.

References

1. Aula, A., Jhaveri, N., and Kaki, M. Information Search and Re-access Strategies of Experienced Web Users. *Proceedings of WWW 2005*, 583-592
2. Bharat, K. SearchPad: Explicit Capture of Search Context to Support Web Search. *Proceedings of WWW 2000*, 493-501
3. Broder, A. A Taxonomy of Web Search. *ACM SIGIR Forum*, 36(2), 2002, 3-10
4. Czerwinski, M., Horvitz, E., and Wilhite, S. A Diary Study of Task Switching and Interruptions. *Proceedings of CHI 2004*, 175-182
5. Iqbal, S. and Horvitz, E. Disruption and Recovery of Computing Tasks: Field Study, Analysis, and Directions. *Proceedings of CHI 2007*, 677-686
6. schraefel, m.c., Zhu, Y., Modjeska, D., Wigdor, D., and Zhao, S. Hunter Gatherer: Interaction Support for the Creation and Management of Within-Web-Page Collections. *Proceedings of WWW 2002*
7. Tauscher, L., and Greenberg, S. Revisitation Patterns in World Wide Web Navigation. *Proceedings of CHI 1997*, 399-406
8. Teevan, J., Adar, E., Jones, R., and Potts, M. History Repeats Itself: Repeat Queries in Yahoo's Query Logs. *Proceedings of SIGIR 2006*, 703-704