IT&SOCIETY, VOLUME 1, ISSUE 3, WINTER 2003, PP. 159-183
http://www.ITandSociety.org

IMPROVING WEB PAGE REVISITATION: ANALYSIS, DESIGN AND EVALUATION

ANDY COCKBURN
SAUL GREENBERG
STEVE JONES
BRUCE MCKENZIE
MICHAEL MOYLE

ABSTRACT

Several years of research suggest improvement is needed in how people return to their previously visited Web pages. Web page revisitation is one of the most frequent actions in computer use, so any interface improvements in this area can have a very large effect. Five categories of revisitation research are involved: 1) Characterizations of user behavior; 2) System models of navigation and their impact on the user's understanding; 3) Interface methods for increasing the efficiency of the Back button; 4) Alternative system models for navigation; and 5) Alternative methods for presenting Web navigation histories. Revisitation is a dominant activity, with an average of 80% of page visits being to previously seen pages. The Back button is heavily used, but poorly understood.

Three interface strategies for improving Web page revisitation are described: 1) A gesture-based mechanism for issuing the frequent Back and Forward commands addresses low-level interface issues; it is shown to be both popular and effective; 2) A 'temporal' behavior for the Back and Forward buttons aims to overcome the problems associated with poor understanding of the current behavior of Back, strongly suggesting that revisitation can be improved by providing temporally ordered lists of previously visited pages; 3) Next-generation browsers could integrate the current tools for revisitation into a single utility, thus allowing simple visualization methods to aid users in identifying miniature target pages.

Andy Cockburn is senior lecturer in the Dept. of Computer Science, University of Canterbury, Christchurch, NZ, where he directs the Canterbury Human-Computer Interaction and Multi-Media Lab. andy@cosc.canterbury.ac.nz.

Saul Greenberg is professor in the Dept. of Computer Science at the University of Calgary, Canada, where he directs the Grouplab Laboratory for HCI and CSCW.

Steve Jones is a senior lecturer in the Dept. of Computer Science at the University of Waikato, Hamilton, NZ.

Bruce McKenzie is a software engineer at the Jade Development Centre in Christchurch, NZ.

Michael Moyle is a software engineer at Allied Telesyn Research in Christchurch, New Zealand.

Every day, millions of people worldwide have problems trying to return to previously visited Web pages. These problems often amount to little more than the minor annoyance of finding that a page has "disappeared'" from those accessible with the Back button. Sometimes, however, extensive searching is necessary to return to a page; for example, users forget to bookmark a "valuable'" page, and they cannot remember how it was originally retrieved. Although the impact of these problems on each individual may be small, it is clear that easing these problems can yield enormous benefits when multiplied across millions of users, and billions of page accesses.

The problems of revisiting Web pages have been examined since 1994, along with designs and evaluations of several systems aimed at improving Web-page revisitation. This article presents an integration and synthesis of this work, in the following order:

1. *User behavior*. Recent results of a Web-use log-analysis show that revisiting pages is a dominant activity on the Web.
2. *System and user models of the current behavior of the Back button.* An easy-to-repeat experiment demonstrates that many users misunderstand the rudimentary behavior of the main interface tool for revisitation—the Back button. Despite this misunderstanding, why is Back heavily used?
3. *Improving the efficiency of the Back command*. The efficiency limitations of the interface mechanisms used to issue the Back command are described, also with an evaluation of a gesture-based shortcut (similar to the scheme recently provided in the Opera Web browser).
4. *Improving understanding and efficiency of the Back model*. An alternative 'temporal' behavior for the Back and Forward buttons is described, with a presentation of the results of its evaluation.
5. *Improving the presentation of revisitation tools*. The implications of the earlier findings demonstrate how the next-generation of Web browsers could integrate and enhance the diverse tools for revisitation that are available in current browsers (Back/Forward, bookmarks, and history lists). This work is ongoing, and preliminary results are encouraging.

USER BEHAVIOR: WHAT DO WEB USERS DO?

Considering that Web browsers are among the most widely used computer applications, there has been only modest research into how they are used. This section briefly summarizes prior analyses of browsing behavior and then describes results from recent studies of what Web users do as they navigate the Web.

Prior analyses: Some researchers detail the demographics of Internet users (e.g., gender percentages, age, occupations, educational attainment, etc.), as well as the demographics of their technologies (e.g., people's connection speed to the Internet and browser selection). The most well-known of these is the 1994-1998 biannual WWW Surveys (http://www.cc.gatech.edu/user_surveys), developed by the Graphics, Visualization and Usability Center at the Georgia Institute of Technology (Kehoe and Pitkow 1996). Recently, sociologists have begun studying Internet use, with a natural emphasis on how Web use changes social structures, rather than focusing on the efficacy of the user interfaces used to navigate the Web (DiMaggio et al. 2001).

Other researchers have analyzed the tasks that people do as they navigate. For example, Byrne et al. (1999) videotaped eight people as they used their browser over the course of their day and codified user behaviors. From this, they developed a task-based taxonomy of browsing, including the six, general Web tasks:

1. Use information—a series of activities in which people use information gathered from the Web,
2. Locate on page—search for particular information on a page,
3. Go to—the act of trying to get the browser to display a particular URL,
4. Provide information—sending information to the browser (e.g., authentication, addresses, search terms),
5. Configure browser—changing the configuration of the browser itself, and
6. React to environment—supplying information or dealing with a problem on demand of the browser.

The authors then proceed to sub-divide these general tasks into more specific ones, and to codify how often they occur. While these results reflect a total of only a few hours of use by a few people, it provides insight into the actual things that people do.

Researchers have analyzed traces (or logs) of users' actions to reveal statistics of use. Some use server-side logs. Since most Web servers collect data indicating when a particular page has been accessed and by what IP address, this is an easy data source to mine. For example, it is relatively simple to analyze logs to expose the frequency of page hits on a Web site. However, server-side logs are limited in that they often do not distinguish well between different users; they collect no data on actual browser use, and they are often missing crucial data. Pirolli, Pitkow, and Rao (1996) and Chi, Pirolli and Pitkow (2000) discuss the problems of extracting meaningful information from server-side logs. Another option is to specially equip the browser so that it logs the users' actions. The advantage of this "client-side

logging" is that it can record the exact history of the user's actions with his or her particular browser.

Perhaps the most well-known, client-side log analyses of Web use are by Catledge and Pitkow (1995) and Tauscher and Greenberg (1997). These studies instrumented the then-popular XMosaic browser to record the pages that users visited and the interface mechanisms used to access them. The participants in both studies were primarily staff, faculty and students in university computing departments. Catledge and Pitkow logged three weeks of use by 107 users in 1995, whereas Tauscher and Greenberg analyzed five to six weeks of use by 23 users in 1995. Catledge and Pitkow revealed that the dominant user interface techniques for visiting pages were clicking on hypertext anchors (52%) and on the Back button (41%). Navigating to pages by typing the URL, clicking Forward, or selecting from 'Bookmarks' were all lightly used, accounting for about 2% each. Tauscher and Greenberg confirmed that link selection and clicking Back are the dominant navigation mechanisms, accounting for approximately 50% and 30% of navigation acts.

Tauscher and Greenberg also analyzed *the recurrence rate* of page visits: "the probability that any URL visited is a repeat of a previous visit, expressed as a percentage." They found that the recurrence rate for the subjects participating in their study was 58%, and a reanalysis of the data from 55 of Catledge and Pitkow's subjects produced a recurrence rate of 61%. This result shows that users had previously seen approximately 60% of pages visited. Tauscher and Greenberg's analysis also reveals that users tend to revisit pages just visited a short while ago, access only a few pages frequently, browse in very small clusters of related pages, and generate only short sequences of repeated URL paths.

In these early studies, people rarely used bookmarks (less than 2% of user actions). However, a later 1996 survey by Abrams et al. (1998) suggested that bookmark use was rising. Some 84% of his respondents had more than eleven bookmarks, indicating that people at least had the intention of returning to key pages. Indeed, Pitkow (1996) reported from a survey of 6619 users that "organizing retrieved information" is one of the top three usability problems of using the Web, reported by 34% of participants.

Our Recent Client-Side Log Analysis of Web Use: While all exemplary studies, the findings of these early studies may not reflect current use of the Web. Things have changed considerably since the mid–90s. Instead of early adopters and technologically savvy 'elite' users, people from virtually all demographic backgrounds use the Web (although not equally). Modern Web browsers have highly polished interfaces with features far beyond those supported by XMosaic. Many Web navigation aids such as search engines and Web directories are now a fundamental part of Web use, yet these tools were either in their infancy or did not exist at the time of the prior studies. The technology itself has changed—broadband connections now give nearly

instantaneous response—a sharp contrast to the "World Wide Wait'" of the early days. The kinds of information accessible on the Web are also completely unlike those of the prior decade. Once an academic repository, the Web is now a massive commercial and populist arena.

METHOD

To update and extend the findings of these previous studies, the Web-browsing activities of 17 Computer Science staff and graduate students were analyzed over a 119-day period from early October 1999 to late January 2000[1]. The use of computer scientists in these evaluation(s) introduces obvious risks of generalizing the results to other communities, but one suspects that computer scientists use the Web for purposes similar to those of most users. Both the Catledge and Pitkow and the Tauscher and Greenberg studies used computer scientists.

The data were gathered through the history and bookmark files that Netscape Navigator (versions 4.5–4.7) maintains. Netscape Navigator was the browser used by the participants in their everyday work, and its user interface features are similar to those of other popular browsers. The browser's history file keeps a list of the URLs the user has visited, the time of the user's last and first visit, the number of visits and the title of each page. The bookmark file stores information about the user's bookmarks and their organization into folders. Copies of these files were obtained through incremental backups that were automatically created every night at the University of Canterbury. To eliminate the chance that participants would modify their behavior due to their awareness that their actions were being logged, participants were asked for permission to retrieve their backup files after the terminating date of the study. That is, participants were asked to have their data mined retrospectively, instead of asking them permission to monitor their future Web uses.

Pages visited per day: The participants made a total of 84,841 page visits, spread across 17,242 different URLs, averaging approximately 42 pages each day. The actual number of pages visited by each user per active day would be substantially higher than this, because the average includes weekends and the Christmas/New Year vacation period during which few participants would have used their browsers at work. Even with this underestimation, these results suggest that people visit almost double the number of pages now than in the mid-90s. That is, the mean of 42 pages per day compares to approximately 14 (Catledge and Pitkow 1995) to 21 pages per day (Tauscher and Greenberg 1997) in the earlier studies.

Percentage of pages that participants had previously seen: The premise behind the Back button, history systems, and bookmarks is that people frequently navigate to pages that they have seen before. As already

DOCKET ALARM

Explore Litigation Insights

Docket Alarm provides insights to develop a more informed litigation strategy and the peace of mind of knowing you're on top of things.

Real-Time Litigation Alerts

Keep your litigation team up-to-date with **real-time alerts** and advanced team management tools built for the enterprise, all while greatly reducing PACER spend.

Our comprehensive service means we can handle Federal, State, and Administrative courts across the country.

Advanced Docket Research

With over 230 million records, Docket Alarm's cloud-native docket research platform finds what other services can't. Coverage includes Federal, State, plus PTAB, TTAB, ITC and NLRB decisions, all in one place.

Identify arguments that have been successful in the past with full text, pinpoint searching. Link to case law cited within any court document via Fastcase.

Analytics At Your Fingertips

Learn what happened the last time a particular judge, opposing counsel or company faced cases similar to yours.

Advanced out-of-the-box PTAB and TTAB analytics are always at your fingertips.

API

Docket Alarm offers a powerful API (application programming interface) to developers that want to integrate case filings into their apps.

LAW FIRMS
Build custom dashboards for your attorneys and clients with live data direct from the court.

Automate many repetitive legal tasks like conflict checks, document management, and marketing.

FINANCIAL INSTITUTIONS
Litigation and bankruptcy checks for companies and debtors.

E-DISCOVERY AND LEGAL VENDORS
Sync your system to PACER to automate legal marketing.