

# Much to Know About History

Eelco Herder<sup>1</sup>, Harald Weinreich<sup>2</sup>, Hartmut Obendorf<sup>2</sup>, and Matthias Mayer<sup>2</sup>

<sup>1</sup> L3S Research Center, Hannover, Germany  
herder@l3s.de

<sup>2</sup> University of Hamburg, Germany  
{weinreich,obendorf,mayer}@informatik.uni-hamburg.de

**Abstract.** Users often revisit pages while browsing the Web, yet little is known on the character of these revisits. In this paper we present an analysis of various revisit activities, based on results from a long-term click-through study. We separate backtracking activities from recurrent behavior, discuss the impact of the use of multiple windows, and show that in particular infrequently reoccurring activities are poorly supported by current history support mechanisms. We conclude with a discussion on design implications for more personalized history support.

## 1 Introduction

Web users frequently return to pages visited before [6]. However, the revisitation tools of current Web browsers still have many known shortcomings. The *back button* is the most important history tool, yet its stack-based behavior is shown to be inefficient and confusing [3]. The temporally or lexically ordered *history list* is hardly used, as its presentation is poor, the filtering options are insufficient, and it requires several user actions to access it [6]. Manual maintenance and organization of *bookmarks* is problematic and time-consuming, which results in overly large, unorganized, and outdated bookmark lists [4].

In this paper we present an analysis of user page revisit behavior, based on results from a long-term client-side Web usage study. The results indicate that users have various different reasons and strategies for revisiting pages. We conclude with several design implications for Web browsers to be more adaptive, and to consider the current requirements and behavior of Web users.

## 2 The Study: Data Collection and Preparation

In Winter 2004, we conducted a long-term client-side study with 25 participants from Germany and the Netherlands [8]. Nineteen participants were male and six female. Their ages ranged from 24 to 52 years (mean: 30.5). Sixteen participants had a background in computer science, nine had different backgrounds. The average time span of the logging periods was 104 days, ranging from 51 to 195 days. The data was collected using an intermediary system based on the framework Scone [7]. The system inserted JavaScript events into every Web page to capture many browser events and parameters. The recorded data included times of

page requests, the browser action that led to the request, the document address, title and size, as well as the time spent on it. After removal of artifacts caused by advertisements, automatic reloads, redirects, and frame sets, 137,272 request actions were left for analysis.

### 3 Results: Categorizing Page Revisits

We found an average *recurrence rate* [6] of 47% ( $\sigma = 11\%$ ); per-subject rates ranged from 20% to 71%. The results confirmed the dominance of revisits to a limited set of highly popular pages, as well as the dominance of revisits to pages visited very recently before [6]. The majority of the top  $n$  most popular pages could be categorized as search engines, news sites, participants' personal or institutional Web sites, and individual interest sites.

The two sets of dominant pages represent two different forms of page revisits: *backtracking* - visits to pages visited before in the same session, and *recurrent behavior* - visits to pages visited before in earlier sessions. Following [2] we used a 25.5 minute time-out mechanism for defining session boundaries.

In order to explore the relations between these two distributions in detail, the page requests were broken down into the following revisit categories:

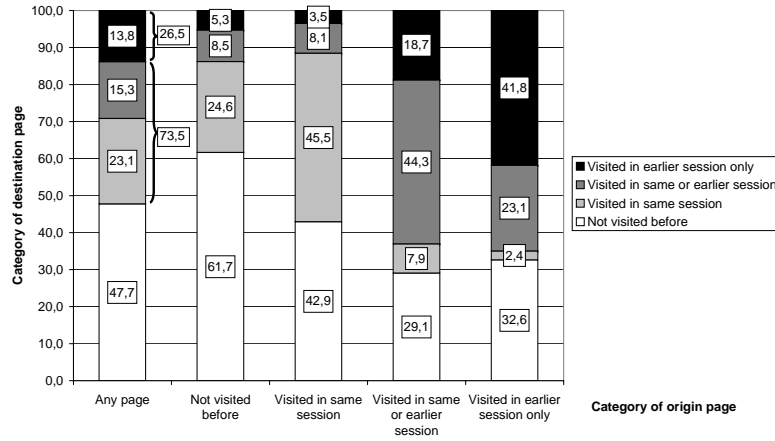
- visits to pages not visited before;
- visits to pages visited before in the same session, but not yet in earlier sessions;
- visits to pages visited before in the same session and in earlier sessions;
- visits to pages only visited before in - one or more - earlier sessions.

The leftmost bar of figure 1 shows the distribution of the page visit categories. 47.7% of all pages are visited only once, backtracking was the most common form of revisitation, covering 73.5% of all revisits; only 26.5% of the revisits involved visits to pages not visited before in the same session.

The second to fifth bar show the transition probabilities between the four categories. An interesting aspect of revisitation behavior emerges: first-time visits, backtracking, and recurrent activities tend to occur in clusters. First-time visits will most likely be followed by another unexplored page. Backtracking activities are mostly followed by another backtracking action, or a first-time page visit. This supports the observation that users frequently backtrack to explore new paths from pages visited before [6]. It can also be observed that in recurrent activities, users backtrack to a similar extent as in first-time visit situations.

#### 3.1 Support for Recent and Frequent Revisits

Page popularity and recently visited pages yield support in different situations: the lists of the top  $n$  most popular pages and the top  $n$  most recently visited pages cover a majority of the pages to be revisited. In order to estimate the performance of the two lists, we calculated the rate of pages that were present in the list of 15 - the number of items presented in the back button pop-up list -



**Fig. 1.** Transition probabilities from one visit category to another

most popular and most recently visited pages for each revisit category (table 1). The results confirm the assumption that the list of the 15 most recently visited pages supports backtracking remarkably well with a coverage of about 90%. As a comparison, we calculated to what extent within-session revisits are covered by the list of pages accessible in the pop-up menu of the back button: the average was only 52% ( $\sigma = 10.8$ ). The list of popular pages supports recurrent behavior with a moderate 72%, mainly due to the long tail of pages that are revisited only a couple of times. However, revisits to pages from earlier sessions that have *not* been visited yet in the current session, are far better supported by the list of most popular pages - 51.2% versus 21.6% - for the 15 most recent pages.

**Table 1.** Revisit support of the 15 most recently visited and the 15 most popular pages

	15 most recent pages		15 most popular pages	
<i>revisit type</i>	<i>average</i>	<i>st.dev</i>	<i>average</i>	<i>st.dev</i>
same session	94.6%	2.97	42.3%	31.05
earlier session	21.6%	9.74	51.2%	13.05
same and earlier	87.9%	6.25	71.6%	14.10

Another drawback of the current back button implementation appeared when we analyzed the application of multiple browser windows and tabs. Our participants tended to use multiple windows to a large extent: 10.5% of all navigation activities involved the opening of a new window or tab [8] - in earlier studies this value was less than one percent [2][6]. The top third group of participants who opened new windows most often, employed the back button to a lesser extent (10.2%) than the bottom third (16.4%); this confirms that multiple windows are

used as an alternative to backtracking ( $t=2.509$ ,  $p < 0.05$ ). A disturbing consequence of this behavior is that it disrupts the concept of the back button: If a user follows trails in multiple windows, the backtracking history is split into separate unrelated stacks. Hence, one often needs to remember what action was performed in which window or tab to relocate a recently visited page.

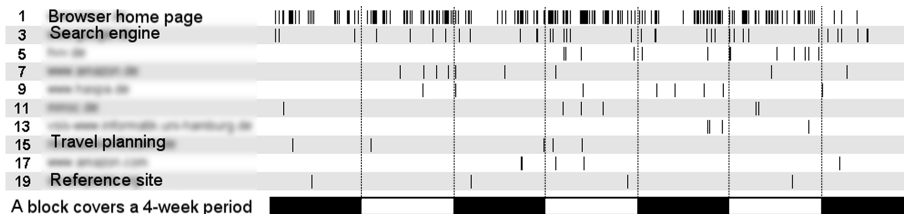
### 3.2 Support for Rare Revisits

Figure 2 illustrates that the average interval between two subsequent revisits is longer for less popular pages; for the pages with a popularity ranking below 10 it was not uncommon that more than a week had passed between two visits.

During an interview we asked every participant about specific situations in which they found it difficult to revisit a page. Most pages that were considered difficult to relocate provided specific information to be reviewed rather than online applications. Situations mentioned by our participants included ‘finding a soccer results list in an unstructured club Web site’, and ‘locating a physician’s home page’. They had either no bookmark, forgot the query or the Web address.

Our results showed that only a few very popular pages are visited on a frequent - often daily - basis. Our participants used different methods to return to these pages: they made use of the bookmark menu or toolbar, or they typed the URI into the address bar, making use of the automatic URL completion function of the browser. Ironically, the longer the interval between two subsequent revisits - and the more likely that the user will not remember the address - the less presumable it is that the action is still present in the browser history. Hence, users had to find other ways to relocate the page; search engines were stated as a key alternative.

We analyzed the revisit category of pages navigated to from search result pages: the far majority (79%) of all page visits following a result page were first-time visits and only 9% were followed by a revisit to a page known from an earlier session. Although only 2% of all queries were entered more than once, 24% of all long-term revisits were preceded by an earlier query. Whereas this suggests that query-based search is a common strategy for relocating pages, our participants stated that they often had problems in remembering the original query for relocating a page and had to resort to wayfinding strategies.



**Fig. 2.** Distribution of revisits to some of the 19 most popular pages of one participant. URIs have been blurred for privacy reasons.

## 4 Design Implications for Future Browsers

The more prominent use of multiple windows and browser tabs requires a major rethinking of the history mechanisms of browsers. A linear history of most recent revisits, as proposed by [6], does not reflect the character of parallel trails, and the unrelated back button stacks do not take the temporal relations between the trails into account. An alternative solution would be a ‘*branching history*’, which shows the trails in temporal order, but separates the activities in different windows. In an earlier laboratory study [5] we found that users who backtrack using links in Web pages often find informations more quickly than users who rely on the back button. Therefore, we think that adaptive hypermedia techniques, such as automatic link annotations, should be considered to visually identify the anchors that point to recently visited pages, in particular those that serve as hubs.

For rare long-term revisits, users often need to rely on finding waypoints that lead to the desired page. Web search appears to be an ineffective manner to find these waypoints, as people seem not to have problems to replicate exact queries. Given the large amount of infrequently visited pages, a manually organized list of bookmarks will be incomplete or too time-consuming to handle. A strategy that is likely to be more effective to support relocating information, would comprise explicit history search, with support for recognizing related earlier queries, and annotated *trails* [1] that users can follow from a waypoint to the desired location. For previously followed trails, shortcuts to the destination pages could save much effort from the user.

The results from our study show that there is much to know about a user’s Web history than is currently used for providing effective support for page revisits. A key challenge for the adaptive hypermedia community is to find effective means to put this knowledge into use.

## References

1. V. Bush. As we may think. *The Transatlantic Monthly*, July, 1945.
2. L.D. Catledge and J.E. Pitkow. Characterizing browsing strategies in the world wide web. *Computer Networks and ISDN Systems*, 27 (6):1065–1073, 1995.
3. A. Cockburn and S. Jones. Which way now? analysing and easing inadequacies in www navigation. *Intl. J. Human-Computer Studies*, 45 (1):105–129, 1996.
4. A. Cockburn and B. McKenzie. What do web users do: An empirical analysis of web use. *Intl. J. Human-Computer Studies*, 54 (6):903–922, 2001.
5. E. Herder and I. Juvina. Discovery of individual user navigation styles. In *Workshop on Individual Differences in Adaptive Hypermedia, AH2004*, pages 40–49, 2004.
6. L. Tauscher and S. Greenberg. How people revisit web pages: Empirical findings and implications for the design of history systems. *Int. Journal Human-Computer Studies*, 47:97–137, 1997.
7. H. Weinreich, V. Buchman, and W. Lamersdorf. Scone: Ein framework zur evaluativen realisierung von erweiterungen des webs. In *KiVS 2003*, pages 31–42, 2003.
8. H. Weinreich, H. Obendorf, E. Herder, and M. Mayer. Off the beaten tracks: Exploring three aspects of web navigation. In *Proc. WWW 2006*, 2006.