Characterizations of User Web Revisit Behavior

Eelco Herder
Group of Human-Media Interaction, Dept. of Computer Science
Twente University, The Netherlands
herder@cs.utwente.nl

Abstract

In this article we update and extend on earlier long-term studies on user's page revisit behavior. Revisits are very common in web navigation, but not as predominant as reported in earlier studies. Backtracking is the most common type of page revisitation and is both used for finding new information and relocating information visited before. Search engines are mainly used for finding new information and users frequently backtrack to result pages. Visits to pages already visited in earlier sessions tend to occur in chunks, but it is not straightforward to create a list of most likely pages that will be revisited. We conclude with a short discussion on design implications for user-adaptive revisitation support.

1 Introduction

While browsing the web, users frequently visit pages already visited before. Earlier studies [Catledge and Pitkow, 1995] [Tauscher and Greenberg, 1997] [Cockburn and McKenzie, 2001] have shown that the majority of page requests involves requests to pages visited before. Of these page revisits, a majority is covered by a small number of very popular pages that are visited far more often than any other page. Another interesting observation is that most page revisits involve pages visited only very recently in the past.

As recurrent behavior is predominant in user navigation actions, web browsers provide various navigation tools to support page revisits. The most well-known and most frequently used tool is the browser's back button, which allows for revisiting pages visited earlier in the same session. Tools for revisiting pages from earlier sessions are bookmarks and the temporarily ordered history list - which is hardly used [Tauscher and Greenberg, 1997]. All of these three revisitation tools have several problems that undermine their usability; perhaps the most serious problem is the lack of integration of these tools [Kaasten and Greenberg, 2001]. Various research efforts on the design of novel revisitation tools are reported in the literature [Tauscher and Greenberg, 1997] [Mayer, 2000] [Milic-Frayling *et al.*, 2004]. This research is typically guided by empirical observations such as the frequencies, distributions and probabilities of various forms of page revisits.

In the past decade three long-term studies have been carried out to characterize users' page revisit behavior [Catledge and Pitkow, 1995] [Tauscher and Greenberg, 1997] [Cockburn and McKenzie, 2001]. Results from the most recent study, carried out by Cockburn and McKenzie, were based on data from late 1999 and might need to be updated to better reflect current usage patterns. Also, the studies did not separate within-session page revisits from cross-session page revisits. As will be explained in more detail in the next section, it is useful to discern these two activities, both for the design and evaluation of novel or adaptive revisitation support.

In this article we update and elaborate on findings from the earlier studies, based on a large amount of web usage data collected from 25 users in 2004 and 2005. In the next section we separate various motivations for revisiting pages and explain how current browsers support these types of revisits. We continue with a brief summary of the process of collecting and preparing the data. In section four we deal with the question whether page revisits really account for 81% of all navigation actions - as reported by [Cockburn and McKenzie, 2001] - is likely to be an overestimation. In section five we explore the well-known power laws of favored pages and recency of page revisits; more specifically, we separate within-session revisits and cross-session revisits and characterize the frequencies with which they occur. In section six we briefly describe the relation between search activities and page revisits. In the second-last section we attempt to exploit these characteristics to create lists of pages that are likely to be revisited. In the last section we discuss the results and design implications for more user-centric revisitation support.

2 Within-Session and Cross-Session Revisitation

As mentioned in the introduction, earlier studies have shown that page revisits are very common in web browsing behavior. [Tauscher and Greenberg, 1997] identified the following main reasons for revisiting pages:

- the information contained by them changes;
- they wish to explore the page further;
- the page has a special purpose;
- they are authoring a page;
- the page is on a path to another revisited page.

This frequently cited list of reasons applies to page revisits in general. There is one important subcategory of page revisits that we would like to point out explicitly: revisits to pages visited before in the same navigation session. As the web is a non-linear medium, users typically navigate in a non-linear manner [McEneaney, 2001]. Within the user navigation paths several pages can be recognized from which multiple alternative paths are initiated. These

pages are generally called *hubs* [Kleinberg, 1999] [Milic-Frayling *et al.*, 2004]. Likely candidates for pages to become hubs are sites' home pages and index pages that serve to navigate users to a number of pages - such as tables of contents and lists of search results [Pirolli *et al.*, 1996].

Whereas recurrent behavior is heavily reported in empirical studies, most theoretical models of web navigation take only forward navigation into account [Pirolli and Card, 1999] [Kitajima *et al.*, 2000]. The CoLiDeS model [Kitajima *et al.*, 2000] regards backtracking as an activity that takes place when forward navigation fails, which does not match the empirical observations mentioned in the previous paragraph. In [Herder, 2004] we proposed a model that separated three different categories of navigation actions:

- *searching*, the process of locating information by issuing queries in a search engine;
- *browsing*, the process of viewing and navigating between web pages;
- *backtracking*, the process of reviwing pages visited before, either for reference or as a starting point for an alternative path.

As motivated by [Teevan *et al.*, 2004], the combination of these three navigation activities can be regarded as *orienteering behavior*, which allows users to specify less of their information need explicitly and provides a context in which to interpret the information found. Results indicate that orienteering is a common strategy for relocating and revisiting information from earlier sessions as well. This finding suggests that cross-session revisits usually appear in chunks, in which the same process of searching, browsing and backtracking can be observed.

In an earlier study [Juvina and Herder, 2005] we found that certain patterns of within-session page revisits indicate that users understand and exploit the site's navigation structure; users that displayed these patterns used within-site navigation support for page revisitation rather than the browser's back button and showed more confidence that they could relocate the pages to be revisited at a later point. Recently, an enhanced implementation of the back button has been proposed that explicitly supports backtracking to hub pages, including sites' home pages, search results and bookmarked pages [Milic-Frayling *et al.*, 2004].

In the remainder of this paper we will further characterize user web revisit behavior and the prevalence of backtracking in both situations in which users look for new information and situations in which users are revisiting information from earlier sessions. In order to provide a context for interpreting the results, we first briefly summarize the process of collecting and preparing the data used in this study.

3 Data Collection and Preparation

The participant pool consisted of 25 participants, of which 17 were recruited from Hamburg, Germany and 8 living in Enschede, the Netherlands. The participants were recruited by personal invitation and were not paid for participation. Nineteen participants were male and six female. Their ages ranged from 24 to 52 years, with an average age of 30.5 years. The majority of the participants worked in the field of computer science, mainly in an academic context. This implies that the results from this study most likely cannot be generalized to *all* users. However, as earlier long-term web usage studies made use of similarly composed participant pools [Catledge and Pitkow, 1995] [Tauscher and Greenberg, 1997] [Cockburn and McKenzie, 2001], we can use these results as a base for comparison.

The participants were logged for some period between August, 2004 and March, 2005. The average time span of the actual logging periods was 104 days, with a minimum of 51 days and a maximum of 195 days. As a comparison, the web logs used by [Cockburn and McKenzie, 2001] consisted of four months of web usage data of 17 participants in total, during which 84.841 page requests were recorded.

The data was collected using a proxy server, which is part of the Java-based framework Scone [Weinreich *et al.*, 2003]. Basic data included the times at which the page requests took place, the web address, title and size of requested pages and the time spent on the pages. Javascript events were inserted into the web pages to capture additional information, such as the opening and closing of browser windows and tabs. As the participants were required to register themselves in the framework and were instructed to turn off their browsers' caching, no heuristics were needed to separate users or to reconstruct navigation paths [Cooley *et al.*, 1999]. For splitting the participants' navigation data into sessions, we used the common 25.5 minutes session time-out heuristic, as established by [Catledge and Pitkow, 1995].

During the logging period 162.605 page requests in total were recorded. A significant number of these page requests turned out not to be initiated by the users themselves but to be automatically generated as a side-effect of user actions. The major categories of these artifacts are:

- *advertisements*, typically embedded into other pages using iframes;
- *automatic reloads*, mainly news sites which refreshed after some time period;
- *automatic redirects*, mainly on dynamically generated web sites;
- *frame sets*, various single files that together constituted one page view.

Various heuristics were used for identifying these artifacts, including server exclusion lists, patterns in web addresses and temporal aspects. As an example, automatic reloads typically generated peaks in the otherwise power law distribution of time spent on web pages. In total 12,1% of the page requests were identified as artifacts, leaving 142.869 page requests that were used for analysis.

4 How Often Do People Revisit Web Pages

Earlier studies have shown that page revisits are very common in web browsing behavior [Tauscher and Greenberg, 1997]. The common formula used for calculating the percentage of revisits among navigation actions is

$$R = 100\% \times \left(1 - \frac{\textit{individual pages visited}}{\textit{total page visits}}\right) \quad (1)$$

Interestingly, the average percentages shown in these studies indicate that page revisits have become more common in the past few years. From the log data from 1994 and 1995 [Catledge and Pitkow, 1995] [Tauscher and Greenberg, 1997] revisit rates of respectively 58% ($\sigma = 9\%$) and 61% ($\sigma = 9\%$) were reported; [Cockburn and McKenzie, 2001] report a significantly higher percentage, 81% ($\sigma = 10\%$), of revisits in their web logs of late 1999. According to [Baldi *et al.*, 2003] this higher percentage might indicate that the usage of the web may have evolved 'from a

more exploratory mode in 1994-1995, to a more utalitarian mode by 1999', a mode where regular visits to sites as general news, travel planners and bulletin boards are predominant. From the 2004 ranking of top 50 websites [comScore Media Metrix, 2004] it can be observed that these sites - entertainment and ecommerce oriented sites in particular - are highly popular indeed compared to 1996.

Based on the above observations, we expected the average percentage of page revisits in our data to be as high or even higher than the 81% reported by Cockburn and McKenzie. Much to our surprise, the average revisit rate for our participants was only 51% ($\sigma = 10\%$). Per-subject revisit rates ranged from a minimum of 26% to a maximum of 79%. As our participant pool is very similar to the participant pools used in the earlier studies, this significant difference either indicates a dramatic change in web usage or it might be caused by differences in the way the data is analyzed.

As an alternative reason for the differences in revisit rates between the studies, [Baldi *et al.*, 2003] suggested that the estimation of the revisit rate from a finite time-window may be an underestimation, as long-term revisits may not be captured during the logging period. However, as indicated in the previous section, the time window used in our study is about the same as the time window used by Cockburn and McKenzie - 104 days on average - and exceeds the time window used in the other studies. We analyzed the effect of the length of the logging period on the page revisit rate and found that the revisit rate stabilizes after about 1000 page views, a number that is reached by each participant in about 10 active logging days, see figure 1.

We found that the reported increase in page revisits is in fact due to differences in data preprocessing. In their report, Cockburn and McKenzie mentioned several data cleaning steps that were taken before analysis. One particular step appeared interesting: URLs involving search queries were truncated to remove suffixes of the form *?name=value&name=value...*, which implies that all queries to search engines, as well as various dynamically generated pages, were generalized into visits to just one page. The authors indicated that this cleaning step did not distort their results and that the characteristics reported were similar to the 'unclean' data. We reanalyzed our data after removal of the query terms, which resulted in a revisit rate of 70,8% ($\sigma = 8,2\%$). This percentage is even higher on the unprocessed data - 73,7% ($\sigma = 8,5\%$), which is the value we used for comparison with the Cockburn and McKenzie study, as they did not report any removal of artifacts caused by adservers, frame sets or automatic reloads. The 95% confidence interval is between 70,2% and 77,2%, which is most likely to have an overlap with the confidence interval on the Cockburn and McKenzie data, which they did not report in their paper.

We leave it up to the reader which of the estimates represents best the actual amount of page revisit behavior. Personally, we prefer the lower estimate as the query part of a url 'serves to identify a resource within the scope of the URI's scheme and naming authority' [Group, 2005]. In practice this means that search engine queries or dynamic page locators often result in different web pages to be loaded. However, from the above it has become clear that the reported increase in page revisits between 1995 and 1999 is most likely due to a difference in data preprocessing rather than a change in general web usage; the average percentage of page revisits has remained fairly stable during the past decade.

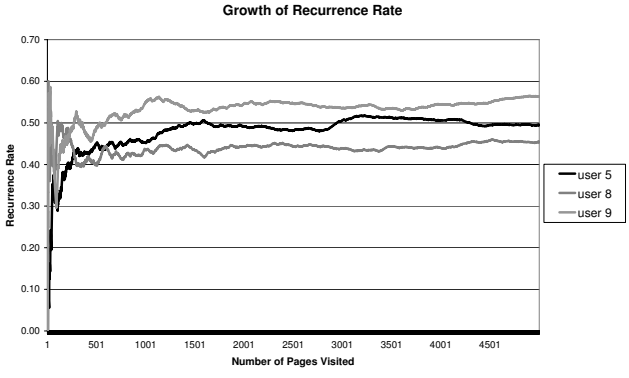

Figure 1: From the picture above it can be observed that the page revisit rate stabilizes after a short period

5 Characterizing Page Revisits

Both [Tauscher and Greenberg, 1997] and [Cockburn and McKenzie, 2001] reported two important distributions of revisited pages. First, there is the *recency effect*; from figure 3 it can be observed that the majority of page revisits involves revisits to pages visited very recently. Second, there is the dominance of favored pages; as can be observed from figure 2, only a small number of frequently revisited pages accounts for the majority of all page revisits.

Although there is likely to be an overlap between these two distributions, they characterize two different forms of page revisits: respectively, visits to pages visited before in earlier sessions and visits to pages visited before in the same session. Based on the two distributions alone it is hard to find out which of the two is the most predominant. In order to explore this in more detail, we identified and annotated the page requests with the following revisit categories:

1. visits to pages not visited before;
2. visits to pages visited before in the same session;
3. visits to pages visited before in one or more earlier sessions;
4. visits to pages visited before in the same session and in earlier sessions.

Following [Tauscher and Greenberg, 1997] we used a 25.5 minute time-out mechanism for detecting session boundaries.

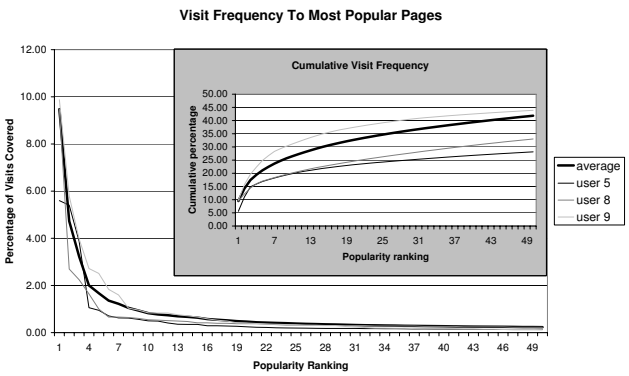

Figure 2: Page popularity rank versus the percentage of revisits covered. The fifteen most popular pages account for about 30% of all revisits

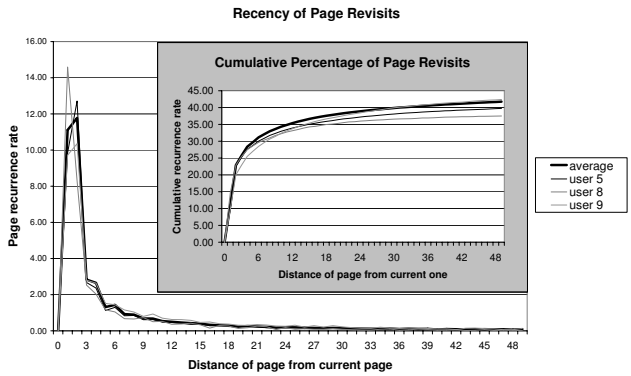

Figure 3: Percentage of page revisits as a function of distance from the current page. The four most recently visited page account for about 50% of all revisits

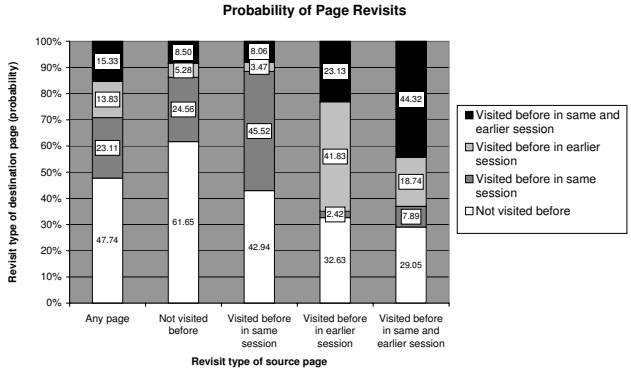

Figure 4: Transition probabilities from one visit category to another. The leftmost bar depicts the overall distribution of categories. It can be observed that users are most likely to remain in the same visit category.

In figure 4 the distributions of page visit categories are displayed. From the leftmost bar, which conveys the overall distribution, it can be observed that within-session page revisits represent the most common form of revisitation, covering 73,54% of all revisits. 44,22% of the page revisits involves revisits to pages not visited before in earlier sessions. Given the predominance of within-session revisits, it is not surprizing that the back button is the most commonly used revisitation tool.

The four remaining bars show another interesting aspect of recurrent behavior: first-time visits, within-session revisits and cross-session revisits tend to occur in chunks. First-time visits are the most common type of visits and if users visit a page for the first time, they will most likely continue to do so. Within-session revisits are likely to be followed by either another within-session revisit or a visit to a new page. This confirms the observation that users frequently backtrack to explore new paths from pages visited before [Pirolli *et al.*, 1996] [Tauscher and Greenberg, 1997] [McEneaney, 2001]. Users who visit pages visited before in earlier sessions will most likely continue to do so as well, backtracking to a similar extent as in first-time visit situations. However, there is still a fair chance of a little more than 30% that they will leave the already visited pages for a page not visited before.

From figure 5 it can be observed that within-session revisits and cross-session revisits may occur at any point in a navigation session. Obviously, the very first page visits are

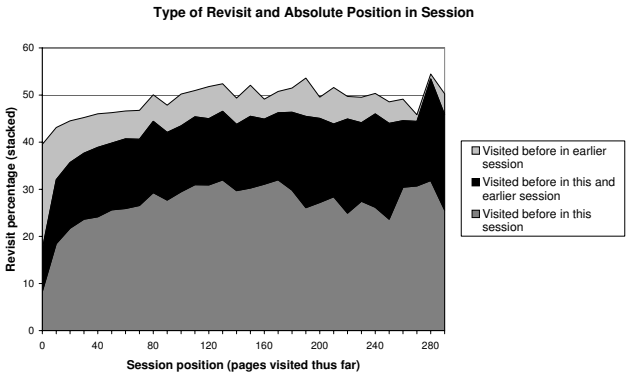

Figure 5: Page revisit probabilities throughout sessions. The spiky behavior in the chart for longer sessions is caused by the low sample rate.

unlikely to be within-session revisits, although the ratio stabilizes pretty quickly. For the same reason, short sessions have relatively few within-session revisits. Interestingly, the amount of revisits in general is a little higher in shorter sessions, mainly due to a large number of cross-session revisits. This confirms the intuition that shorter sessions are more utilitarian in character than longer sessions.

6 Revisits and Search

From the preceding section it has become clear that both within-session revisits and cross-session revisits are important aspects of user web navigation. Obvious questions that come to mind is to what extent search engines are used to relocate information and to what extent recurrent behavior is part of searching. In order to find this out, we annotated all visits to the search engines' result pages. Visits to the search engines' home pages were not considered, as these were used as the browser's home page by some of our participants; in addition, some participants used other means to issue queries, such as a browser-integrated toolbar.

All our participants used Google as their main search engine. In the web logs we found 16868 visits to Google result pages. On average, 12,27% ($\sigma = 6,55\%$) of the participants' page visits were visits to Google result pages. This number includes results for new queries, requests for additional results, query modifications and returning to result lists.

We considered the revisit category of pages navigated to from these result pages as the main indicator as to what extent search engines were used for relocating pages. The far majority, 79%, of all page visits folowing a result page were first-time visits. A modest amount of 9% search results were followed by revisit to a page visited in earlier sessions but not yet in the current session - in 24% of the cases the query used had been used in earlier sessions as well. The remaining 12% of the search results were followed by a within-session revisit, either or not to a page already visited in an earlier session. From these numbers it becomes clear that searching is used to relocate information, but to a modest degree; finding new information is the main activity that search engines are used for.

The above observation is confirmed by the correlations between the number of revisits and the extent of searching in navigation sessions. The percentage of page revisits is significantly lower in sessions with many visits to result pages (r=-0,359; $p < 0,01$). Interestingly, the percentage of within-session revisits is higher in search-intensive ses-

sions (r=0,133; $p < 0,01$). As indicated in section 2, this is most likely due to users returning to hubs in order to follow an alternative path. These hubs are mainly the result pages themselves, as 30,86% of all visits to Google involved returns to the result page - a smaller percentage of 26,67% were the results of a new query.

7 Supporting the Next Page Revisit

From the characterizations in the preceding sections it appears to be fairly straightforward to create a set of pages that a user is most likely to revisit next. In this section we construct some basic models for constructing lists of likely candidates for page revisitation, based on the identified power law distributions. A probabilistic model was chosen in order to include candidate pages that were not in the list of fifteen most popular or most recent pages but that still had a fair chance of being revisited. For each user the web log was randomly split into a training set and a test set, each containing half of the registered sessions. We evaluated performance for each page visit category ten times for each user.

The first model makes use of the popularity distribution of pages, which is expected to be useful for supporting cross-session revisits. For each visit in the test set fifteen pages were selected from the popularity distribution. Results for each page revisit category are listed in table 1.

revisit type	*average*	*st.dev*
same session	6,63%	4,01
earlier session	31,11%	11,33
same and earlier	45,77%	17,28

Table 1: Performance of Popularity-Based Model

From the table it can be observed that the popularity distribution does not perform too well in predicting which page will be visited next. By contrast, a simple list of fifteen most popular pages would have been helpful in 51% of all cross-session revisits and 71% of visits to pages visited in same and earlier sessions. This can be explained by the long tail of pages that are revisited only a couple of times; in addition, we found that the most popular pages were news sites and search engines, pages that are likely to be revisited on a very regular basis and more than once in the same session. For these reasons, the probabilistic approach did more harm than good.

The second model makes use of the recency effect of page revisits, which is expected to be useful for supporting within-session revisits. For each visit fifteen pages were selected from the user's recent history; the probability of a page to be selected is according to its distance from the current page, given the user's recency distribution. Results for each page revisit category are listed in table 2.

revisit type	*average*	*st.dev*
same session	72,72%	5,55
earlier session	25,71%	11,76
same and earlier	78,77%	8,68

Table 2: Performance of Recency-Based Model

According to the expectations, the recency-based model performs fairly well for within-session revisits; in seven out of ten cases the next page visited was included in the set of possible candidates. Prediction of cross-session revisits is slightly higher than the popularity-based model, although this difference is not significant. The list of fifteen most recent pages would have been helpful to about the same extent as our popular model; whereas the within-session revisits would have been supported in 94% of the cases, the performance for cross-session revisits is slightly lower. This indicates that it would be useful to support revisits to recent pages beyond a distance of fifteen pages; most likely, an additional emphasis on recognized hub pages [Pirolli *et al.*, 1996] [Milic-Frayling *et al.*, 2004] will further increase the performance.

As a comparison, we ran a simulation to find out to what extent within-session revisits are covered by the list of pages behind the back button. As explained by [Cockburn and Jones, 1996], the list of pages behind the back button is organized as a *stack*. The number of pages that can be directly accessed via the back button is fifteen in most browsers. When using the back button, the user moves down in the stack; the forward button can be used to go to the top of the stack. If a user decides to make use of other navigation tools than the back and forward buttons, all pages that are located above the current location are removed from the stack. We simulated all navigation sessions in the web logs and reconstructed the contents of the back button stack for each page request. We assumed that each session started with an empty stack and filled or popped the stack with each page request. The average percentage of within-session revisits that was supported by the back button for each user was 51,7% ($\sigma = 10,8$). This confirms the observation of [Tauscher and Greenberg, 1997] that the stack-based approach may be suitable for very short-distance revisits, but is outperformed by a recency-ordered history list.

8 Discussion

In this article various aspects of user web revisit behavior have been dealt with. Although recurrent behavior is less predominant than reported in earlier studies, it still comprises 51% of all navigation actions. The majority of page revisits is accounted for by within-session revisits, many of which are backtracking activities. In current browsers backtracking activities are mainly supported by the back button, which is surprizing as problems associated with the stack-based model and the unexpected removal of pages from the top of the stack while using the back button have been reported in various papers [Cockburn and Jones, 1996] [Mayer, 2000].

In addition to browsing and backtracking, search is an important type of web navigation. Search engines are mainly used for finding new information. When looking for new information, users frequently backtrack to the result list in order to initiate an alternative path. Whereas most browsers allow users to open a result page in a separate window or tab, this requires an additional effort that most users appear not to take. As browsing, searching and backtracking are very interrelated activities when looking for new information, support for these activities should be integrated. An enhanced back button that explicitly supports backtracking to hub pages and search results, as proposed by [Milic-Frayling *et al.*, 2004] is a good start. However, it does not provide necessary meta-information on the search context [Teevan *et al.*, 2004].

Contrary to earlier observations [Cockburn and McKenzie, 2001] [Baldi *et al.*, 2003] revisiting pages from earlier sessions plays a relatively modest role in user web naviga-

tion. However, once users are revisiting 'old' information, they will most likely continue to do so. Although the distribution of popular pages would suggest that a list of *n* most popular popular might be effective for supporting cross-session revisits, this turns out to be effective in only 51% of all cases. This might be an explanation why bookmarks are still the most important type of support for cross-session revisits and why users generally fail to construct and manage their bookmark collections [Cockburn and McKenzie, 2001].

As revisits to 'old' pages tend to appear in chunks, with backtracking as prevalent as when visiting 'new' pages, it makes sense to invest effort in developing and evaluating algorithms for recognizing hubs that are frequently reused to provide effective starting points for orienteering in previously accessed parts of the web. As users rarely repeat search engine queries, candidate pages are most likely to be found in other sites than search engines. Even though users are reported not to repeat long trails, it is most likely still worthwhile to annotate earlier trails in order to support repeated tasks. Earlier laboratory research [Juvina and Herder, 2005] indicated that users benefit from these task-related annotations, as they provide additional context without suggesting the user that they should be blindly followed.

More research on the design of effective, context-aware revisitation support is needed. In order to create effective support mechanisms, insight in general revisitation patterns is needed. In this article we presented several characterizations of user revisit behavior and several design implications. In particular the distinction between within-session revisits and cross-session revisits appears to be an effective starting point for rethinking and evaluating future revisitation support.

Acknowledgments

The navigation data discussed in this paper is collected in the context of a long-term study in collaboration with Matthias Mayer, Hartmut Obendorf and Harald Weinreich from the VSIS department of the University of Hamburg. The author's work is sponsored by the Dutch Innovative Research Program IOP-MMI.

References

[Baldi *et al.*, 2003] P. Baldi, P. Frasconi, and P. Smyth. *Modeling the Internet and the Web*. John Wiley & Sons, 2003.

[Catledge and Pitkow, 1995] 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.

[Cockburn and Jones, 1996] 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.

[Cockburn and McKenzie, 2001] 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.

[comScore Media Metrix, 2004] comScore Media Metrix. Surfing down memory lane to january 1996: comscore media metrix revisits first-ever web site rankings, February 25 2004.

[Cooley *et al.*, 1999] R. Cooley, B. Mobasher, and J. Srivastava. Data preparation for mining world wide web browsing patterns. *Knowledge and Information Systems*, 1 (1), 1999.

[Group, 2005] Network Working Group. Uniform resource identifier (uri): Generic syntax, January 2005.

[Herder, 2004] E. Herder. Sniffing around for providing navigation assistance. In *Proc. German Workshop on Adaptivity and User Modeling in Interactive Systems ABIS2004*, 2004.

[Juvina and Herder, 2005] I. Juvina and E. Herder. The impact of link suggestions on user navigation and user perception. In *Proc. User Modeling 2005*, 2005.

[Kaasten and Greenberg, 2001] S. Kaasten and S. Greenberg. Integrating back, history and bookmarks in web browsers. In *Proc. Chi 2001*, pages 379–380, 2001.

[Kitajima *et al.*, 2000] M. Kitajima, M.H. Blackmon, and P.G. Polson. A comprehension-based model of web navigation and its application to web usability analysis. *People and Computers*, XIV:357–375, 2000.

[Kleinberg, 1999] J.M. Kleinberg. Authoritative sources in a hyperlinked environment. *Journal of the ACM*, 46 (5):604–632, 1999.

[Mayer, 2000] M. Mayer. Contextual web visualization: Browsingicons and browsinggraphs to support orientation and navigation in the world wide web. In *Proc. 6. Tagung der Deutschen Sektion der Internationalen Gesellschaft fr Wissensorganisation*, pages 267–280, 2000.

[McEneaney, 2001] J.E. McEneaney. Graphic and numerical methods to assess navigation in hypertext. *Intl. J. Human-Computer Studies*, 55:761–786, 2001.

[Milic-Frayling *et al.*, 2004] N. Milic-Frayling, R. Jones, K. Rodden, G. Smyth, A. Blackwell, and R. Sommerer. Smartback: Supporting users in back navigation. In *Proc. WWW 2004*, 2004.

[Pirolli and Card, 1999] P. Pirolli and S.K. Card. Information foraging. *Psychological Review*, 105 (1):58–82, 1999.

[Pirolli *et al.*, 1996] P. Pirolli, J. Pitkow, and R. Rao. Silk from a sow's ear: Extracting usable structures from the web. In *Proc. CHI 1996*, pages 118–125, 1996.

[Tauscher and Greenberg, 1997] L. Tauscher and S. Greenberg. How people revisit web pages: Empirical findings and implications for the design of history systems. *Intl. J. Human-Computer Studies*, 47:97–137, 1997.

[Teevan *et al.*, 2004] J. Teevan, C. Alvarado, M.S. Ackerman, and D.R. Karger. The perfect search engine is not enough: A study of orienteering behavior in directed search. In *Proc. Chi 2004*, pages 415–422, 2004.

[Weinreich *et al.*, 2003] H. Weinreich, V. Buchman, and W. Lamersdorf. Scone: Ein framework zur evaluativen realisierung von erweiterungen des webs. In *Tagungsband Kommunikation in Verteilten Systemen - KiVS 2003*, pages 31–42, 2003.