

At the Right Time: when to sort web history and bookmarks

Alan Dix

Lancaster University
Computing Department
Lancaster, LA1 4YR
UK

alan@hcibook.com

Jason Marshall

Loughborough University
Sports Technology Research Group
Loughborough, LE11 3TU
UK

J.M.Marshall@lboro.ac.uk

<http://www.hcibook.com/alan/papers/HCI2003-history/>

Abstract

In open-ended interviews web users preferred text interfaces for sorting bookmarks to more complex graphical 2D and 3D interfaces. They also expressed a desire to sort web pages as they bookmarked them. However, in an experimental study we found that recall performance of web pages sorted during browsing was significantly poorer than performance when they sorted bookmarked pages at the end of a browsing session. This effect appeared to decay in a retest a week later. This work shows that users are able to articulate meta-knowledge strategy, but it questions whether users' expressed preferences are a good guide for design.

1 Introduction

This paper describes an empirical study of the effects of different sorting times for web bookmarks and history: incrementally 'during' browsing or all together 'after' browsing.

Studies by Tauscher and Greenberg (1997) have shown considerable revisiting of the same web page during browsing. Some of this is clearly due to backing up after following mistaken links or hub-and-spoke behaviour; indeed Catledge and Pitkow (1995) found that 30% of all navigation is the use of the 'back' button. However, there are a considerable residual number of pages that are 'really' revisited because the user wants to see the content again. Browsers support this behaviour both for short-term revisitation (back button and visit stack) and for the long term (history, bookmarks, favourites). In a formal analysis of several hypertext and web browsers Dix and Mancini (1997) found that the history and back mechanisms were subtly different in them all. This emphasises results found in other studies that users find back and history confusing and this is reflected in behaviour with comparatively little use of history or multi-step back. Bookmarks are more heavily used, but still are known to have many problems. There have been some more radical interfaces proposed and used at an experimental level including the data mountain (Roberston et al., 1998), which allows users to arrange thumbnails of bookmarked pages in a 2D landscape, and the WebView interface unifying history and bookmarks (Cockburn et al., 2002).

In order to investigate these issues we performed a short study in two phases. The first phase consisted of a series of exploratory interviews aimed at understanding users views on different proposed visualisation techniques for history and bookmark mechanisms. This was followed by a more formal experimental phase. Our original intention was to perform the more detailed work on

2D and 3D interfaces, but our initial study showed that the most critical issue for users was not how history was displayed, but *when* classification was performed. So, the more formal experiments during the second phase was focused instead on comparing outcomes when users sorted during or after browsing.

2 Phase I – what the users want

Our initial study was focused on a small number of open-ended exploratory interviews to discover typical browsing patterns, understanding and use of bookmarks and history. In addition, interviewees were presented with a number of 2D and 3D visualisation techniques proposed in the literature including data mountain (Robertson et al., 1998),, WebView (Cockburn et al., 2002). and WebBook (Card Robertson & York, 1996). It should be noted that in contrast to Tauscher and Greenberg (1997) and Catledge and Pitkow (1995) numerical studies of web data, this was about users self-reported behaviour and attitudes. There were six interviewees, two female and four male, with ages from 19 to 31 years. All were experienced web users.

The users reported hub-and-spoke browsing behaviour using either search engines or directly typed URLs to start their browsing. Surprisingly when asked what features were most useful for remembering a site four out of the six said that the URL or some derivative of it was most useful. None mentioned actual page content and only one mentioned keywords. This may be a reflection of the web expertise of the interviewees, or may be a reflection of existing history interviews, however, it does call into question interface suggestions that rely heavily on recall of page images and thumbnails. In fact, the users here may be being more knowledgeable than those designing novel interfaces as empirical studies by Kaasten, Greenberg and Edwards (2002) found that thumbnails have to be quite large before they become useful, especially if one is interested in identifying pages rather than simply sites.

When offered images (not live use) of the various proposed history interfaces it was the text based 2D interfaces (Apple Hot List and WebView) that were most popular. Again this should be interpreted with care as they are more familiar looking and also included explicit annotation which most interviewees felt was more important than the particular visual aspects of the history mechanism. This latter point also supports Amento et al. (1999) study where all the participants wanted to record comments about sites as they visited and collected them.

The interviewees were given the opportunity to comment freely on web history and favourites mechanisms and to suggest what they would like to see including drawing sketches etc. Although the main reason for the interviews was to investigate history visualisation, the issue that repeatedly arose was the interviewees' desire to be able to classify bookmarks/favourites at the moment they were 'remembered' rather than as a secondary exercise. Currently all common browsers force a bookmark-now, sort-later mode of working. The strength of the interviewees' reactions led to a refocusing of our empirical studies towards understanding this 'when to classify' issue in more depth.

3 Phase II – what the users need?

Based on the results of phase I study, we decided to focus our more detailed experiments on the issue of sorting during browsing compared with sorting afterwards and its effect on later (online) recall.

Although the interviewees had expressed a desire for 'during' sorting, we postulated that this would in fact lead to less clear classification. This is because when items are sorted during the process of browsing it is not clear the full range of future pages that will require classification, whereas when classifying pages after navigation it is easier to produce a balance and sensible classification. For example, if the first few pages seen are about aspects of football should one classify them as 'football' pages or 'sport' pages. If the following pages include one further page on golf and many on completely different topics then 'sport' would have been the best classification. However, if the rest of the pages were about golf, rugby, cricket etc., then 'football' would have been best. Classification 'after' is able to take this into account.

This therefore lead to a hypothesis that sorting during browsing would be less 'good' than sorting afterwards.

The main condition was the *during* vs. *after* sorting. In both conditions the participants used the Internet Explorer favourites mechanism to classify 50 web pages. In the *during* condition the participants were asked to look at each page and then immediately save it as a favourite and classify it. In the *after* condition the participant would look at all the pages saving each as a favourite and then after all had been seen open the sorter and classify all 50 pages.

For the experiment ten paid participants were recruited, five male and five female, aged between 18 and 36 years. All were experienced users of Internet Explorer and of its favourites mechanism. Two datasets of 50 web pages were selected from 100hot.com's 2002 list of top web sites. The experiment consisted of a within subjects design where each participant performed both a *during* and *after* sorting. The experiment was balanced for order of presentation and for the site set. So half the participants had set 1 with the *during* condition and set 2 with the *after* condition and half the other way round. The quality of participants' classification was judged by asking them to use the classified bookmarks to answer a series of questions. In each case this was immediately after the relevant sorting. Because different sets of pages were used there was no learning effect for content and in fact the results showed no measurable order effect.

Our hypothesis was indeed borne out by the results, which did show significantly ($P < 0.04$) better recall for the *after* condition compared with *during* sorting (see table 1). Also in a post-test questionnaire the participants preferred the *after* sorting, in direct contrast to the interviewees' imagined preference. Other results included a correlation between time spent sorting and performance.

Table 1: Mean retrieval times for sorting method and retrieval cue (in seconds)

	Mean	Std. Deviation
During browsing	584.2	232.73
After browsing	421.8	106.77

A small number of participants were retested a week later. The number was very small (only four participants) and so any results are merely suggestive. However it did appear that the advantage of *during* vs. *after* sorting disappeared almost completely. If the quality of the *during* classification were indeed worse then one would have expected to have had even worse results on retesting after the immediate memory of the classification process had faded.

As the numbers were too small for statistical testing these counter-intuitive results may well be just a random effect. However, they have made us question whether the strength of the *after* sorting may be partly explained by the fact that the sorting process occurs closer to the post-test. More sophisticated experiments may be required to separate all the potential causes.

4 Reflection

There are several lessons from this study beyond the raw results of the experiments.

First the interviewees were able to articulate desires relating to meta-knowledge issues – the timing of bookmark classification. This ability to reflect on as well as engage in knowledge discovery has been investigated further in Dix, Howes & Payne (2003).

Second, the actual running of the experiment also showed how complex these issues are, especially because we are looking at relatively long-term effects that are hard to capture fully within a laboratory setting. We deliberately chose an experimental set-up that was at least partially ecologically valid rather than a more controlled and specific pure psychological experiment. This allowed us to find some real and strong effects, but by its nature admits multiple interpretations.

Thinking about the differences between the experimental condition and ‘real life’ several things are obvious. In a real situation the pages visited would be ones that held some personal interest for the users and hence fit within existing mental structures. In contrast our participants were faced with pre-selected pages. However, the chosen pages were of broad interest, which we hope reduced this effect. Harder to control was the interface itself. Using an off the shelf and familiar interface rather than a do-it-yourself one means that we avoid the frequent situation of testing a ropey prototype. The downside is that the IE favourite mechanism is designed for periodic *after* sorting, which could have biased against the *during* condition. Finally, real use would not fall neatly into one of these camps. Instead on-the-fly sorting would happen when the user had an existing set of categories and so some of the problems of on-the-fly sorting with a blank slate would only occur during the earliest browsing. Also even if on-the-fly sorting were the norm it is likely that heavy users would perform some periodic tidying up of categories.

Finally, the two phases of this study show, what everyone in HCI knows, but we are often reluctant to admit: users don’t always know best. Although the initial interviewees were heavily in favour of on-the-fly sorting, this preference was reversed when faced with doing it in practice. As noted, this latter preference may have been due to a poor interface, but this would be equally worrying if bad experiences influence user requirements so heavily. This was an experimental study, not an exercise in participatory design, but does emphasise that good participatory design should be user focused and may be user led, but always requires strong expert guidance and aid.

5 Acknowledgements

This work was supported by EPSRC under Masters Training Package GR/N63277 and the EQUATOR Interdisciplinary Research Collaboration (www.equator.ac.uk). A complete description of the results in this paper and further results from the same experiments can be found in Marshall (2002).

6 References

- Amento, B., Hill, W., Terveen, L., Hix, D. & Ju, P. (1999). An Empirical Evaluation of User Interfaces for Topic Management of Web Sites. In *Proceedings of CHI 99*, 552-559.
- Card, S.K., Robertson, G.G., and York, W. (1996). "The WebBook and the WebForager: An Information Workspace for the World-Wide Web" in *Proceedings of ACM SIGCHI '96*, Vancouver, Canada, 111-117, April 1996.
- Catledge, L., & Pitkow, J. (1995). Characterising Browsing Strategies in the World-Wide Web. In *Proceedings of the 3rd International World Wide Web Conference, Darmstadt, Germany*, Published in *Computer Networks and ISDN*, 27, 1065–1073, Elsevier Science. <http://www.igd.fhg.de/www/www95/papers/>
- Cockburn, A., Greenberg, S., McKenzie, B., Jasonsmith, M., and Kaasten, S. (1999). WebView: A Graphical Aid for Revisiting Web Pages. *OzCHI'99: Australian Conference on Computer-Human Interaction*, Wagga Wagga, November 28–30, 1999. pages 15-22.
- Dix, A., & Mancini, R. (1997). Specifying history and backtracking mechanisms. In P. Palanque and F. Paternó (Eds.), *Formal Methods in Human-Computer Interaction* (pp. 1–24). London, Springer-Verlag. <http://www.hcibook.com/alan/papers/histchap97/>
- Dix, A., Howes, A., & Payne, S. (2003). Post-web cognition: evolving knowledge strategies for global information environments *International Journal of Web Engineering Technology*, 1 (1) (in press).
- Kaasten, S. and Greenberg, S., & Edwards, C. (2002). How People Recognize Previously Seen WWW Pages from Titles, URLs and Thumbnails. In X. Faulkner, J. Finlay, F. Detienne (Eds) *People and Computers Volume XXVI* (pp. 247–265), Springer Verlag
- Marshall, J. (2002). An Exploratory Study in World Wide Web Navigation History. MRes Dissertation. Lancaster University, UK
available from this paper's web page: www.hcibook.com/alan/papers/HCI2003-history/
- Robertson, G., Czerwinski, M., Larson, K., Robbins, D., Thiel, D., & Dantzich, M. (1998). Data Mountain: Using spatial memory for document management. In *Proceedings of UIST'98* (pp. 153–162). San Francisco, California, November 1998, ACM Press..
- Tauscher, L., & Greenberg, S. (1997). How people revisit web pages: empirical findings and implications for the design of history systems. *International Journal of Human Computer Studies*, 47 (1), 97–138.