A History-Centric Approach for Enhancing Web Browsing Experiences

Yoshinari Shirai
NTT Communication Science Laboratories, NTT Corporation
2-4, Hikaridai, Seika-cho, Soraku-gun, Kyoto, 619-0237, Japan
&
Research Center for Advanced Science and Technology, University of Tokyo
4-6-1 Komaba, Meguro, Tokyo, 153-8904, Japan
way@cslab.kecl.ntt.co.jp

Kumiyo Nakakoji
Research Center for Advanced Science and Technology, University of Tokyo
4-6-1 Komaba, Meguro, Tokyo, 153-8904, Japan
&
NTT Communication Science Laboratories, NTT Corporation
2-4, Hikaridai, Seika-cho, Soraku-gun, Kyoto, 619-0237, Japan
kumiyo@kid.rcast.u-tokyo.ac.jp

Yasuhiro Yamamoto
Research Center for Advanced Science and Technology, University of Tokyo
4-6-1 Komaba, Meguro, Tokyo, 153-8904, Japan
yxy@kid.rcast.u-tokyo.ac.jp

CHI 2006, April 22–27, 2006, Montréal, Québec, Canada.
ACM 1-59593-298-4/06/0004.

Abstract
Browsing Web pages, which plays an important part of our daily creative knowledge work, often includes purposefully revisiting pages we have browsed before. Many of the existing tools and approaches for revisitation clearly distinguish the use of history from Web browsing. The approach presented in this paper blurs the distinction between browsing the Web and visiting stored pages in a personal Web browsing history database. The HCB (History-Centric Browsing) system allows a user to browse a previously visited Web page stored in the history database in the same way as browsing a page on the Web. The system associates pages from the database to the currently displayed page through three types of relevancies: temporal sequence, URL/location-based proximity, and content similarity. The HCB-stat, HCB-vis, and HCB-tempo components use the associations to enrich the current page-viewing experience.

Keywords
Personal Web browsing history, enhanced Web browsing experience, interaction design

ACM Classification Keywords
H5.m. Information interfaces and presentation (e.g., HCI): Miscellaneous

Introduction

Browsing Web pages plays an important part of our daily creative knowledge work. In so doing, we not only browse pages that are "new" to us but also purposefully revisit pages we have browsed before.

Such revisitation of information has been found to be both widespread and important. Cockburn et al. have presented empirical evidence that more than 80% of information uses are for revisitation [3].

Mechanisms such as bookmarks, back and forward buttons, and histories have been integrated with Web browsers to help users revisit Web pages. WebView [4] is an extended Web browser with the enriched "back" button with a list of thumbnail images of previously visited pages. Browsing Icons [8] displays animated graphs of users' browsing paths as they surf the Web. The Global Tree browser [10] and Domain Tree browser [5] visually present trees of a personal Web browsing history adjacent to the currently viewed page. MindRetrieve [9] and browseback [2] are browsers specifically built to help users search in the history of personal Web browsing by using keywords and/or thumbnail images of pages. TimeScape visually displays the history of local files along a time line [11]. GoogleDesktop [6] finds local files as well as previously viewed Web pages based on keywords. Google Search History [7] stores a user's search queries and helps him/her revisit pages visited at the time of each search.

Such existing tools and approaches, however, primarily treat Web browsing and the use of history as two separate activities. Claiming that revisitation should be regarded as an essential part of a Web browsing activity, this paper presents a history-centric approach to enhance Web browsing experiences. The approach seamlessly integrates browsing the Web and visiting stored pages in a personal Web browsing history database.

We have implemented a mechanism for Microsoft Internet Explorer that collects a Web browsing history and stores it in the HCB (History-Centric Browsing) history database. When a user opens a Web page, either by directly typing a URL, by selecting a bookmark, or by clicking on a link in a different Web page or a document, the HCB database system takes a snapshot of the opened Web page, produces a bitmap image of the browser window, stores the source HTML code together with the subcomponent files, records the URL and its Web page title, and marks the time stamp.

The HCB system allows a user to browse a previously visited Web page stored in the history database in the same way as browsing a page on the Web. The system associates pages from the database to the currently displayed page through three types of relevancies: temporal sequence, URL/location-based proximity, and content similarity. The HCB-stat, HCB-vis, and HCB-tempo components use the associations to enrich the current page-viewing experience.

Revisitation Needs

Revisitation of a Web page is not merely browsing a previously revisited page. Questions a user may ask regarding revisitation include:

- What was the page I was just looking at before this page?
- Have I already bookmarked this page?

- What was the page that had a content similar to the currently viewed page?
- Has anything changed in this page since I visited it the last time?
- Have I seen this page before?
- What was the occasion/context for which I have previously seen this page?

These questions almost always arise in the context of the current page-viewing task. In fact, revisitation, or more broadly, the use of a personal Web browsing history, becomes necessary to support the current viewing task; in other words, this history is not used merely for the sake of understanding the history per se.

The history-centric approach we describe in this paper views browsing a page, both from the Web or from the history database, as a single type of a viewing task. When the user displays a page in the browser, regardless of whether the page is from the Web or from the database, the currently displayed page becomes a part of the user's browsing history.

The HCB System

Based on this framework, the HCB system associates pages stored in the database to the currently displayed page, and presents different types of information about them in terms of the currently displayed page through the three components, HCB-stat, HCB-vis, and HCB-tempo (Figure 1).

Associations

The HCB system uses three different types of associations.

– *Associated pages by temporal sequence* are those viewed immediately before and after the currently displayed page (from the history database).

– *Associated pages by URL/location-based proximity* refer to different versions of pages that have the same URL as the currently viewed page. To identify versions of a single URL, we first identify entries in the database that have the same URL, and then use an HTML hash table to identify identical contents and reduce the number of redundant contents, resulting in versions that have the same URL but different contents.

– *Associated pages by content similarity* are those that have page contents similar to the currently viewed page. We use a simple keyword-matching algorithm among the titles of Web pages, but could employ more sophisticated algorithms, such as those described in [1].

HCB-stat

Figure 1(a) shows how HCB-stat presents a user with the statistical summary of the pages associated with the currently displayed page by using the shapes of the three icons. HCB-stat minimizes interference with the current viewing experience; five seconds after a user opens the Web page, HCB-stat gradually emerges in the top part of the currently displayed window.

The bar-like icon on the left indicates where the time stamp of the current temporary page lies within the history. The top half of the center icon indicates the number of the newer versions and the bottom half indicates the number of the older versions of the currently viewed page existing in the database. The

density of the icon on the right indicates the number by content similarity.

HCB-vis

HCB-vis overlays the visualized summary of the associated pages on top of the currently displayed page. Figure 1(b) shows a sample screen. The center column shows the URL and the time stamp of the currently displayed page.

Three sets of thumbnail images of Web pages are displayed in three lines. The vertical line shows thumbnail images of different versions of the Web page of the current page. They are temporally ordered with the older ones toward the bottom. The horizontal line shows thumbnail images of pages successively viewed when the current page was browsed in the past. Five linearly ordered windows on the left are those visited immediately before the current page, and those on the right are those visited immediately after it. The linearly ordered thumbnail images in the bottom of the grayed area display Web pages that have Web page titles similar to the currently displayed page. We eliminate any versions that have the same URL as the current page from this list.

Hovering a mouse cursor over one of the thumbnail images presents a larger image of the page together with its time stamp and URL, and clicking on the image opens the page in the browser.

HCB-tempo

In support of exploring the personal Web browsing history, we use the Search Query-based Slicing (SQS) technique [12] to index a potentially large volume of Web page history. A search query posted in Google represents a user's partial intention when originally viewing the page. Our premise is that the search result page partially shares the context with the subsequently viewed pages. The use of search queries makes sense to the user because the target space of search and exploration is a personal Web browsing history. Looking at search queries may remind the user of the time when the query was posted. Segmented parts "sliced" by a sequence of search queries, therefore, may provide a larger granularity than pages for exploring the personal Web browsing history.

To identify search queries, the HCB system parses all the Web page titles, and extracts the keyword field from the Web page title of each Google search result page.

Figure 1(c) shows HCB-tempo, which allows a user to explore the space of the personal Web browsing history in terms of the currently displayed page. The right-most column shows a list of thumbnail images of the sequentially ordered history pages. A red-bordered thumbnail represents the currently displayed page. The left-most column shows a list of search queries. The search query posted immediately before the currently displayed page is highlighted. The two columns in the middle represent the time lines. The left time line shows the entire time of the history database, and the right column is a zoomed-in view for the search query session.

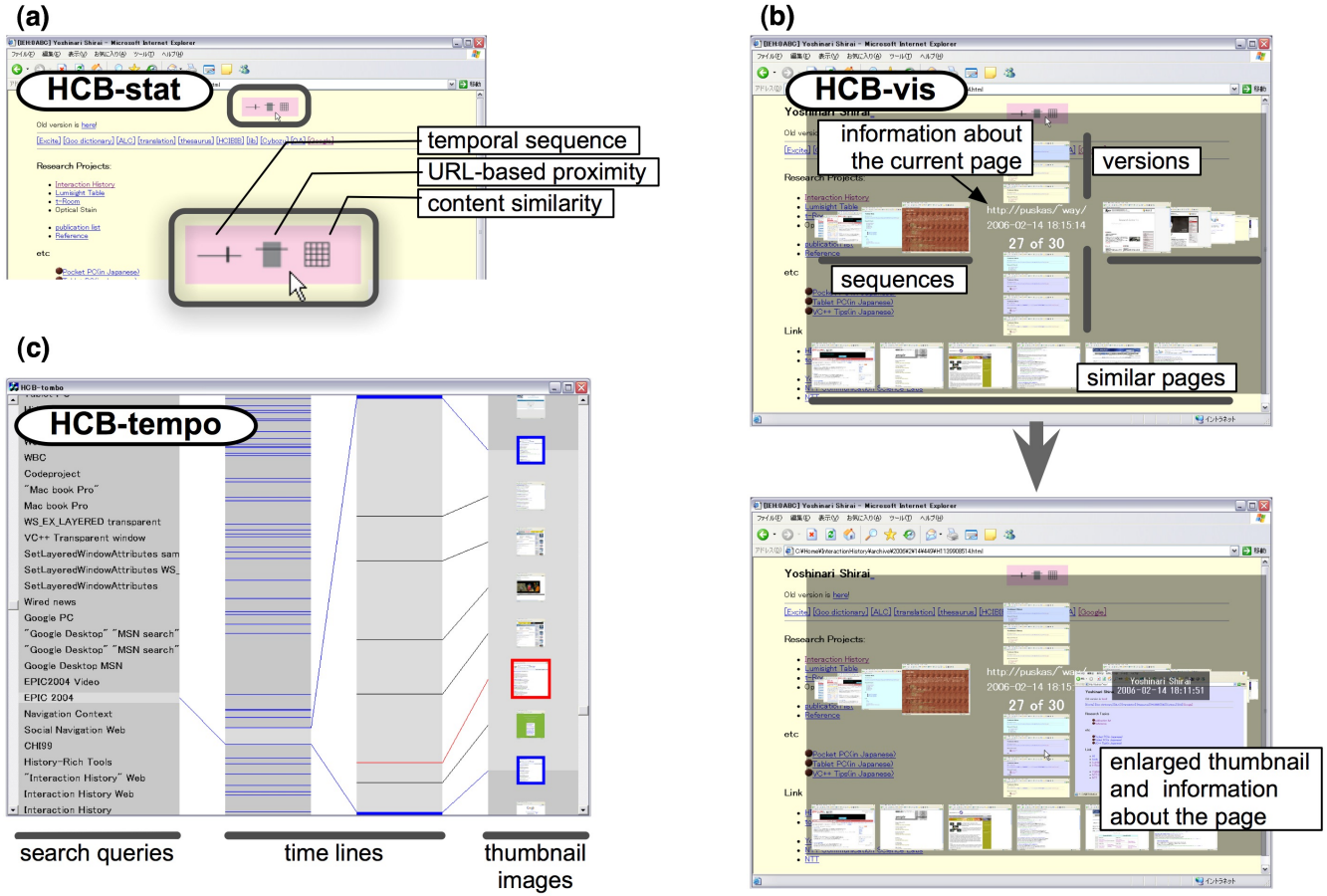

figure 1. HCB-stat, HCB-vis and HCB-tempo

Discussion

From a perspective that revisitation of the information space is important, we present our history-centric approach to support Web browsing. Our goal in this research is not to help the user to understand the Web browsing history per se, but rather to help the user in the "current" Web browsing experience, enhanced with the existence of the personal Web browsing history. Our future work is to observe users using the HCB system in their browsing tasks.

Acknowledgments

This research is partially supported by the Ministry of Education, Science, Sports and Culture, Grant-in-Aid for Scientific Research (A), 16200008, 2004-2007. We would like to thank Yunwen Ye at L3D Center, University of Colorado at Boulder, for valuable discussions on earlier versions of this research.

References

[1] Baeza-Yates, R. and Ribeiro-Neto, B., Modern Information Retrieval, Reading, MA. Addison-Wesley, 1999.

[2] browseback, http://www.smileonmymac.com/browseback/

[3] Cockburn, A. and McKenzie, B., What Do Web Users Do? An Empirical Analysis of Web Use, International Journal of Human-Computer Studies, Vol. 54, No. 6, pp. 903-922, Academic Press, 2001.

[4] Cockburn, A., Greenberg, S., McKenzie, B., Jasonsmith, M., and Kaasten, S., WebView, A Graphical Aid for Revisiting Web pages, Proceedings of the OZCHI'99 Australian Conference on Human Computer Interaction, 1999.

[5] Gandhi, R., Kumar, G., Bederson, B., and Shneiderman, B., Domain Name Based Visualization of Web Histories in a Zoomable User Interface, DEXA Workshop, pp. 591-600, 2000.

[6] GoogleDesktop, http://desktop.google.com/

[7] Google Search History, http://www.google.com/searchhistory/

[8] Mayer, M. and Bederson, B. B., Browsing Icons: A Task-Based Approach for Visual Web History, HCIL Technical Report, 2001.

[9] Mindretrieve, http://www.mindretrieve.net/

[10] Nadeem, T. and Killam, B., A Study of Browser History Mechanisms for Web Navigation, 5th International Conference on Information Visualization, 2001.

[11] Rekimoto, J., Time-Machine Computing: A Time-centric Approach for the Information Environment, ACM UIST'99, 1999.

[12] Shirai, Y., Yamamoto, Y., and Nakakoji, K., Search Query-based Slicing for a Personal Web Browsing History, Advanced Visual Interfaces 2006, Short Paper in Extended Abstract, May 2006 (submitted).