Contextual Web History: Using Visual and Contextual Cues to Improve Web Browser History

Sungjoon Steve Won, Jing Jin, Jason I. Hong

Human-Computer Interaction Institute

Carnegie Mellon University

{swon, jingj}@andrew.cmu.edu, jasonh@cs.cmu.edu

ABSTRACT

While most modern web browsers offer history functionality, few people use it to revisit previously viewed web pages. In this paper, we present the design and evaluation of Contextual Web History (CWH), a novel browser history implementation which improves the visibility of the history feature and helps people find previously visited web pages. We present the results of a formative user study to understand what factors helped people in finding past web pages. From this, we developed CWH to be more visible to users, and supported search, browsing, thumbnails, and metadata. Combined, these relatively simple features outperformed Mozilla Firefox 3's built-in browser history function, and greatly reduced the time and effort required to find and revisit a web page.

Category and Subject Descriptors

H.5.2 User Interfaces: User-centered Design

General Terms

Design, human factors

Keywords

Web history, web browser, revisitation, re-finding

INTRODUCTION

A revisitation occurs when a user returns to a web page that has been previously visited. One previous study indicated revisiting web pages makes up 58% of all Internet browsing [26]. Another study suggests that revisitation is common, with 81% of web pages having been previously visited [12]. However, web browser features intended to aid revisitation are rarely used. In particular, Weinreich et al found that web browser history was used to initiate only 0.2% of all page requests [29]. Other work has found similar results [5, 8, 16, 17], which is surprising given that all modern web browsers come with history built in.

In formative user studies that we conducted (presented in this paper), we found two reasons why browser history was rarely used. First, the history is typically a "hidden" feature that is not as prominent in the browser window as the Back button or address bar. As a result, as in past work [5, 16, 29], we found that many people were not even aware that history exists. Second, history implementations in popular web browsers only provide rudimentary support for browsing and searching previously visited web pages. For example, by default, Microsoft Internet Explorer 7.0 only shows the page titles of the visited sites. Similarly, Mozilla Firefox 3 only shows the page titles on the history sidebar, and the page titles and URLs in the Library window (accessed by clicking "Show All History"). Both browsers organize the history by titles, dates visited, and frequency of visits, but past work suggests that this format is not very useful for users [21]. In short, current history implementations are hard to find, and provide few cues for helping people find pages they have seen in the past.

In this paper, we present the design, implementation, and evaluation of Contextual Web History (CWH), a history-tracking component whose goal is to improve the usability and utility of the history feature in web browsers. We first conducted formative studies to understand why people didn't use the history function, and what cues people found useful for recall and recognition of infrequently visited web pages. To address the lack of awareness of history functionality, we integrated CWH with existing browser and search engine functionality, the latter of which is a variant of work by Teevan [27]. For example, when entering search terms into popular search engines, CWH also provides relevant information about past web pages. To address the difficulties in using existing web histories, CWH provides a richer set of cues about the content of a page, including time of visit, visual appearance, and text search. We evaluated CWH against the browser's default history and other methods of revisitation (such as search
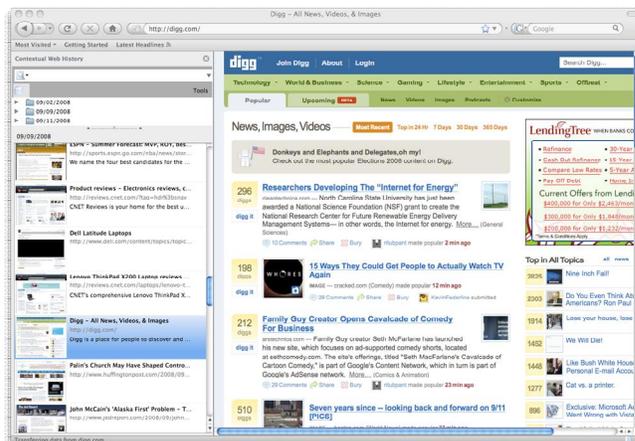

Figure 1: CWH prototype. On the left are thumbnails of sites matching the current query. On the right is the selected page.

engines), and showed significant improvements over both.

Through multiple iterations of paper prototypes, interactive prototypes, and user studies, we found a good combination of several relatively straightforward ideas for supporting revisitation, including helping people find the history feature, browse and search through the history, see relevant information about past pages, and quickly find previously visited web pages again.

RELATED WORK

Many projects have studied recording, organizing, and revisiting information on the web. We have organized past work into studies on revisitation patterns, support for revisitation (those explicitly built into web browsers as well as those that are not), and improving web browser history.

Studies on Revisitation Patterns

There has been much work on revisitation patterns [4, 9, 23, 26]. Revisitation can be broken down into four categories based on time [23], specifically short-term (within an hour, 72.6% of all revisits), medium-term (within a day, 12% of all revisits), a week (7.8% of all revisits), and long-term (greater than 1 week, 7.6% of all revisits). In this taxonomy, our goal with CWH is to support medium and long-term revisits, sometimes called "re-finding" information. Adar et al [4] present a large scale analysis of search logs combined with a survey of volunteers, bringing to light several clusters of revisitation patterns. Our work was informed by these papers, and focuses on building a better tool to facilitate revisitations.

Current Revisitation Methods

Web browsers are equipped with a variety of tools to assist revisitation, such as auto-completion of URLs, the back button, bookmarks, and history. However, these tools only aid revisitation in limited situations. For example, auto-completion is only applicable for commonly used URLs that were often already known by users [23]. The back button is useful for short-term revisitations, however it is unable to support long-term revisitation over multiple browsing sessions [13, 22].

There has also been a great deal of work examining the use of bookmarks to revisit past web pages [3, 12, 16, 17]. Bookmarks allow users to specify and save links to what they anticipate revisiting. However, they were rarely used in practice [12, 16]. Reasons for low usage of bookmarks include bookmarks being concealed in folders, too many bookmarks, and too much effort to organize bookmarks. Another problem with bookmarks is that they require users to proactively record a web page for future revisits. In contrast, our goal is to design a system that supports revisiting any web sites from history.

To compensate for the lack of support for long-term revisitation, people also used search engines to re-find information [17, 23, 28]. Our formative user studies confirmed this behavior. Similarly, Aula et al [5] conducted a survey showing that search engines were the most commonly used technique that was not directly supported by web browsers. However, one common problem they found was that users could not often remember what keywords were used, and was thus a highly iterative process. With CWH, users can search their local history rather than searching the web at large. Thumbnails and contextual information are also provided to aid the process.

The Re:Search Engine [27] is a system that enhances existing search engines to help people find as well as re-find information. When a search query is made, it fetches new search results while simultaneously fetching memorable results from the previous searches for the same query. The Re:Search Engine is useful when the target page was originally found through a search engine. Like the Re:Search Engine, we found that integrating web history with search was a useful way to expose history functionality. However, in contrast, our work with CWH seeks to improve the browser history functionality in general, which is applicable regardless of how a person got to that page. Furthermore, a person can use CWH directly if they know that they are looking for a specific page they have already viewed, rather than also seeing search results for pages that may not be relevant for their current needs.

Improving Web Browser Back, Bookmarks, and History

Several techniques have been developed to help improve revisitation. For example, Firefox 3 comes with a URL locator bar with better search capabilities, such as using snippets of URLs, snippets of site titles, and user-generated tags in bookmarks. Past versions could only search bookmarks and do substring matching on browsing history.

LeeTiernan et al [21] built and evaluated two different web history modifications, the first clustering individual entries by web site and time, the second clustering web pages by similarity. Both approaches were shown to represent the user's mental model better than the default Internet Explorer history. To better accommodate users, Kaasten and Greenberg [19] proposed improving the back button by integrating Back, History, and Bookmarks together.

Other projects have experimented with different ways to improve the presentation of web browser history. A common approach is to include thumbnail images of web sites. MosaicG [6] and PadPrints [15] both displayed web page thumbnails in a graph that represented browsing history. A PadPrints study found that the number of erroneous page access and the time needed to find the correct page both decreased dramatically. WebView [11] shows page thumbnails in a hub-and-spoke or in a temporal scheme. Kaasten and Greenberg [19] also provided a thumbnail for each visited page, which could also be bookmarked. Wexelblat developed Footprints [30], an information system that made use of real-world metaphors

such maps, signposts, and wear and tear, to expose history of use by a group of users.

There have also been commercial and open source products to improve web history. For example, BrowseBack [1] is a commercial product that lets you search your web history by keyword and also displays thumbnails of web pages. MindRetrieve [2] is an open source desktop search engine that indexes the text of all web pages you have seen. It presents search results as a rank ordered list.

Cockburn and Greenberg [10] describe two issues that revisitation systems have to address: page representation and page organization. Using their taxonomy, we use a combination of text labels and images of rendered pages, while somewhat supporting a temporal scheme (coupled with search filtering, an issue not discussed in that paper).

Our work with CWH builds on the work above, and differs in that we are interested in developing and evaluating a holistic design that integrates a number of features to facilitate re-finding web pages. This includes searching and browsing, providing multiple cues, and also making the history functionality easier to find in the first place. Another goal was to increase usage of history functionality without being intrusive. We aimed for a solution that is integrated with the browser's style and is simple to use.

FORMATIVE USER STUDIES
We first conducted several formative user studies on browser history, to gain a qualitative understanding of how people found previously visited web sites and what aspects of those web sites were most useful and memorable for retrieval. This user study consisted of an interview and several recall and recognition tasks.

There have been several studies testing people's ability to recall attributes of documents, in the context of document retrieval [14, 18, 20, 24]. Much of this past work relies on novel spatial metaphors or spatial memory to aid recall. We chose not to pursue this path, instead relying more on a combination of existing searching and browsing style interfaces that people are more familiar with and are easier to build on top of the existing web browser APIs.

Blanc-Brude and Scapin studied what people recall about their documents, in the context of desktop PCs [7]. They found that people were best at recalling file location, file type, associated events (e.g., emails, phone calls), and visual elements. This study helped inform us in the design of our studies and CWH user interface. However, file type is not as meaningful on the web, since the large majority is HTML, and file location may be a better aid for personal files on personal computers, but less so for the web. Visual elements were useful for us, and were incorporated into CWH in the form of thumbnails to aid recognition.

We recruited seventeen participants using online listings and fliers. Participants were initially told that the study was about web browsing and were compensated $20 USD at the end of the study. Web history and re-finding were not mentioned until later in the study. Participants ranged in age from 18 to 55, most being under 35. All accessed the Internet daily and approximately 80% used Firefox as their primary browser. While our participants heavily favored Firefox, we believe that our results are still valid given the similarity among web browsers in their history features. Furthermore, all of our Firefox participants had previous experience using IE or Safari. Participants spent approximately 15 to 20 hours per week using web browsers.

Interviews
We started with an interview to understand participants' revisitation patterns and their understanding of the default history features in browsers. Participants were asked to recount the last time they revisited a website when they did not remember the URL. They were then asked how they would generally approach such a situation. While a very small percentage of the participants said they would try to go to a web site that linked to the destination page (that is, re-tracing their steps as to how they got to the desired page in the first place [23]), over 95% of our participants said that they would use a search engine.

We also wanted to confirm participants' understanding of the default browser history feature. When asked about how frequently they used the browser history, 57% of our participants didn't know such a feature existed, and 43% did not use it due to its lack of visibility. Only two of our 17 participants (part of the 43%) had used history within the past month, but they preferred not to use it. Three participants said that they recalled incidents in which they used the history to look for and revisit a web page but were not successful due to too many unhelpful results.

Recall and Recognition Tasks
In the recall and recognition portion of the study, we conducted user studies to see what aspects of web pages people could easily remember. Our goal here was to gain a qualitative understanding of what aspects of web pages people could remember, enough to inform the design of CWH, rather than conducting an in-depth study of human memory with respect to web pages.

Preparations for the Recall and Recognition Tasks
One day prior to the study, participants were asked to visit web pages selected by the researchers. Initial participants were asked to visit four pages while later participants were asked to visit six pages. We selected the web pages using StumbleUpon, a Firefox extension for finding interesting sites. We selected pages that would be of interest to a general audience, such as news pages and image galleries. Participants were asked to spend at least five minutes per page and then email the researchers after they finished.

Part One: Recall Tasks

Potential Memory Cue	%participants who recalled cue about page
Colors on a site, excluding images	100%
Visual structure & layout of page	94%
Time user visited the site	83%
Logos, prominent images	82%
Presence of animated content	58%
Title of the page	31%
Domain name of the URL	17%
Path and Filename of URL	17%

Table 1. Recall for aspects of a web page for 17 participants.

Participants were asked to recall two web pages randomly selected by the researchers from the set of web pages that they were asked to visit before the study, as well as the last web page they visited before the study. Researchers referred to the pages by general descriptions such as "the page about a basketball team" or "a shopping site". The other web sites that people were asked to visit were used later in the recognition tasks below.

Participants were first asked to list any characteristic of the web page they could remember in a freeform manner. They were then asked to examine a list of potential memory cues that we created (Table 1), to see if they could remember any more details about the web page. We created the list of potential memory cues based on characteristics of a web page that we thought would be easy to record in a web browser and potentially useful for helping people recall web pages. We verified correctness of the descriptions provided by the participant by visiting the described website.

During the free recall task, over 88% of the participants focused on textual content, for example, a summary of key points of a news article. This was not surprising since the most common method to look for a web page is to enter keywords into a search engine. However, when presented with the list of potential memory cues in Table 1, many participants were able to recall many other characteristics of those web pages. Below is breakdown the various cues, ordered by percent of participants who remembered that cue for a particular page.

Colors on a site: This cue describes the general color scheme of a web page, such as the color of the background or the main content area. All participants were able to remember colors well. Note that color codes used in HTML and CSS are easy to extract, but colors within images are harder to detect. Also, having people match colors by specifying RGB or HSV values may be difficult in practice, so approximate matches may be more useful here. Therefore, we considered approximate descriptions as correct, such as describing a red, yellow, and purple background as "warm, with red and other vibrant colors."

Visual structure & layout of a page: This cue describes the overall layout of major web page components, such as header, sidebar, body, and footer. Participants were most accurate with the location of the title, navigation bar, and main content area of web pages. Some participants were also able to sketch out the general layout of a web page.

Time user visited the site: This cue describes the approximate hour of the day the participant visited a web page. This cue was marked accurate if the participant knew what time of the day he or she visited the web page, such as "yesterday afternoon". Many participants remembered the times at which they visited web sites by relating to activities they performed before and after visited web sites, a form of episodic memory. For example, a participant might remember working on a class paper before visiting a website to find more information. The exact time at which a web site is visited is already recorded by the web browser, and mapping exact times to descriptive times is possible. Note that our results are somewhat biased because people were asked to visit sites the day before, making them more likely to remember them than if it were, for example, a week before. However, this timeframe of a day before is still within the scope of revisitation we want to support.

Logos, prominent images: This cue describes logos, icons, or other images that can be associated with a web page. Participants were able to easily recall large images or images relevant to the text.

Presence of animated content: This cue includes any animated content, such as flash animation. Participants could remember the presence of animated content on web pages quite well; however, this cue was not found to be generally useful because much of the animation was in the form of advertisements, which were present on many websites and thus not helpful in the recall tasks.

Title of the page: This cue describes the title of a web page in the content area of a web page, as opposed to the HTML title tag, since these two do not always match. Some participants were able to recall the entire title or phrases within the title that were unique to the page. These titles can be retrieved by making the assumption that the largest text on the page is the title. However, some web sites substitute text titles with title images, so titles cannot always be reliably extracted from web pages.

Domain name portion of URL: This cue describes the domain name of a web page. For example, if a web page's address is "http://www.site.com/page.html," its domain would be "site" or "site.com" Memorable domains were short and related to the page's content.

Path and Filename portion of URL: This cue describes the part of a URL after the domain name. For example, with "http://www.site.com/abc/page.html," the path would be "abc" and the filename would be "page.html". Most participants were not able to remember any part of the URL unless it was short and descriptive of the page it pertained to. An example of a memorable part of a URL would be "maps" from "http://www.google.com/maps/…".

In general, we found that participants could recall high level text and visual characteristics fairly well, including such things as the general page content, visual appearance, and kinds of images. Examples include summary of a news article, color scheme of a web page, and an image of a basketball game. Also, the amount of detail participants provided was often quite accurate and useful in recognizing the actual web site when we visited the site afterwards.

Note that this study is limited in that the sites were visited only the day before, but as stated earlier, it is within the scope of revisitation that we want to support, and still provides some useful information on what cues were more useful for participants. Another limitation of this part of the study is the context in which our participants visited the sites. More specifically, they visited the sites by clicking on links in our email, rather than through a search engine or through a link on a blog. Thus, the path by which a person got to a page is a potentially useful cue that we could not readily capture in this portion of the study.

Part Two: Recognition Tasks
The goal of this part of the study was to evaluate how well thumbnails would aid in revisiting web pages, and what size was best for participants. There has been much past work using thumbnails [6, 11, 15, 19, 24] and on augmenting thumbnails [25, 31]. Our goal was not to innovate directly on thumbnails, but rather understand the best way of integrating and using them for our CWH prototypes.

Participants were asked to identify the thumbnails of the web pages they visited among eight to twelve thumbnails (Figure 2), which contained thumbnails of web pages they visited and did not visit. As stated earlier, initial participants were asked to visit four web pages, and performed the recognition task with eight thumbnails. Later participants were asked to visit six web pages and performed the recognition task with twelve thumbnails.

Participants were asked to select the thumbnails of web pages that they had visited for the study and explain aloud why those thumbnails corresponded to web pages they visited (i.e., with the visited web pages that were not used in the recall task). We started this process with thumbnails of size 85x55 pixels, and then enlarged them to 143x89 and then 235x148. We selected the dimensions of smallest thumbnail because the size made it hard to distinguish layout and images in a web page. The medium and large size thumbnails are each 1.6 times the previous size. At each step, participants could confirm or reject their previous choices. For the early participants, four of the eight thumbnails were distracters. For later participants, eight of the 12 thumbnails were distracters.

Success in identifying thumbnails varied across the three different sizes. The number of errors participants made on the small and medium thumbnails was similar. However, for the largest thumbnails, each participant averaged less

Figure 2. Thumbnails of actual pages participants visited prior to the interview, along with distracters. These were shown to participants to evaluate effectiveness of visual aids in recall task.

than one error in identifying all of the appropriate thumbnails. The participants cited large text or images, the page's layout, and the colors as clues in the correct thumbnail, especially for the largest thumbnails. This is consistent with the most memorable features we identified in the recall task. In cases where the participant identified a thumbnail incorrectly, she either corrected herself while explaining the reason to the experimenter, or identified the correct one after being told that the thumbnail originally chosen was not one of the web pages she visited.

Recap of Key Findings and Preview of CWH Prototypes
Half our participants were not aware of the browser's history functionality. Other users stopped using the history feature because it provided too many results and the information displayed was not helpful. To address the former issue, our eventual solution was to integrate CWH with the browser search functionality, and to embed relevant historical results within web-based search engine results. To address the latter, CWH offers multiple ways of quickly filtering and sorting history results. Initially, we only supported searching on history, but later found that browsing also useful for finding recent web pages.

Through user studies, we found that people were best at recalling text and visual characteristics of web pages. Other characteristics, such as time of visit and URL were somewhat useful, so we opted to include these features, giving users multiple ways to search and filter results.

CONTEXTUAL WEB HISTORY PROTOTYPES
Informed by the formative user studies above, we iterated on several paper prototypes and two interactive prototypes built as Firefox extensions. We will focus on the interactive prototypes and defer description of the implementation until the second prototype. Participants in our evaluation studies were recruited using fliers and online postings.

FIRST PROTOTYPE AND EVALUATION
In our design, we added a search toolbar and a toolbar icon to the web browser chrome to increase visibility of CWH (Figure 3). The search toolbar lets the user access CWH by providing a search query. Users could input different types of information such as keywords, dates, and times. All input was processed within the single textbox so that the user could quickly get to the results. For example, if the user enters "meatball pasta yesterday morning," CWH would search for web pages containing the words "meatball" and "pasta" that were accessed yesterday morning, which by default is defined as between 8am and 11am. Our data

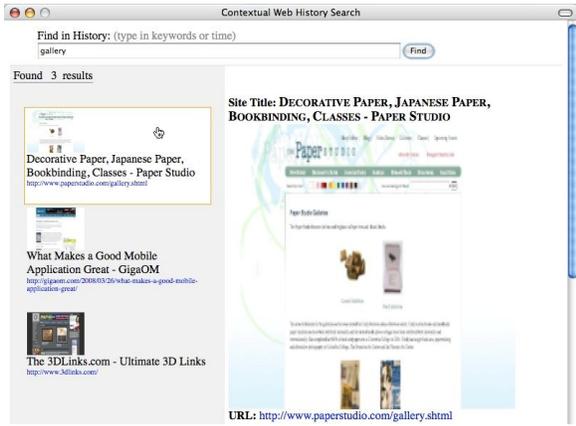**Figure 3. The icon and search field on the toolbar improve visibility of the history feature.**

Figure 4. The main window of the first CWH prototype. At the top is the search field. The list of relevant search results is on the left, with the highlighted thumbnail being the selected item. On the right is detailed information on the selected item.

showed that adding the search bar did succeed in increasing visibility, with 75% of our participants using this feature to access history. Later, in our second prototype, we integrated the search toolbar with the Firefox 3's search bar located on the top right of the web browser. Another possibility would be to integrate it with the URL locator bar, though we found it difficult to implement and thus did not pursue this option.

The main history window composed of a search bar at the top, a result list on the left, and a details area on the right (Figure 4). We displayed information about page title, URL, thumbnail, and keywords (as defined by the HTML meta tag). We also showed how a result matched the search query. We also showed a list of recently visited web pages if a query has not been entered. We originally added this feature so that people would not be presented with a blank screen when they opened CWH. However, we found showing these recently visited pages provided another method of revisiting web pages and made it much easier to find recent pages. The research literature also suggests that recently visited web pages are likely to be revisited in the near future [9, 23]. This also supports Adar et al's [4] idea of supporting both a working stack (recently accessed pages) and a searchable stack of infrequent pages.

We conducted two evaluations of our first prototype. The first evaluation compared the effectiveness of our first prototype against both the Firefox 3 history and the participant's preferred method of revisitation in a laboratory setting. The second evaluation compared the effectiveness of CWH against the default browser history in a more realistic study that takes into account the web surfing habits

of the user. Participants were paid $15 USD.

First Evaluation of First Prototype

We evaluated CWH with 12 participants in a between-subjects study, comparing (1) our CWH prototype, (2) Firefox 3's default browser history, and (3) the participant's preferred method of revisitation. Participants were between 20 and 50 years old, and were mostly Firefox users.

Participants were e-mailed two lists of web sites to visit prior to meeting with the researchers. The first contained four sites and was sent six days before the experiment. The second contained three sites and was sent three days before experiment. Participants were asked to email us to confirm when they had finished visiting the sites on the lists.

To provide a uniform experience, participants used one of our computers that had a web history pre-loaded with the correct relative dates. This history included the sites that we asked participants to visit, and other sites similar in topic, content, or appearance. In total, there were 40 web pages in history, 20 pages being six days old and 20 pages being three days old. Participants were also not told that this was a study regarding web history until the day of the study.

Each participant was asked to find a website using only one of the three methods mentioned above: CWH prototype, Firefox 3 Browser History (the "Show All History" feature), and the participant's preferred method. Each site was cued in the same way as the recall task. Each method was evaluated with four participants. Participants were asked to look for three web pages from three days ago, and three web pages from six days ago.

The effectiveness of each method was evaluated by counting the number of actions the user performed. For total number of actions, we summed up the number of intentional mouse hovers and clicks, words typed, and directional keystrokes.

Key Findings
When asked to use their preferred method to revisit a web page, participants either navigated to the desired page from a related page, or performed a search on a search engine.

A common problem we saw with Firefox's default browser history was in narrowing down results. For example, finding a web page that was in the "older than x days" folder of the browser history would lead to too many pages for the user to browse. This resulted in user actions such as repeated scrolling and multiple attempts at filtering using different search text inputs. Because of this, participants found CWH's feature to search using specific dates useful.

A second problem participants encountered was that many URLs had no contextual meaning to the user (e.g. session IDs). Third problem was that Firefox's history displays only a web page's title, leading to problems of repeated titles as shown in Figure 5. CWH's thumbnail and snippet
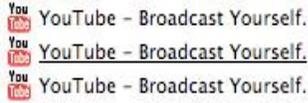
Figure 5. Repeated web page titles in Firefox's history.

of web page's text content were useful in distinguishing sites like these without actually revisiting each of them.

In revisiting web pages from three days ago, our prototype was 25% more efficient than Firefox's history and 31% more efficient than the user's preferred method. In revisiting web pages that were visited six days ago, our prototype was 51% more efficient than Firefox's history and 29% more efficient that the user's preferred method (Figure 6). The difference of ratios in the 3-day and 6-day conditions resulted because when participants were looking for sites in the 6-day category, they were less precise about when they visited the site, therefore they browsed through more entries in history than in the 3-day category, resulting in an increase in number of steps and time. Overall, participants took less steps looking for a web page using CWH compared to other methods.

Second Evaluation of First Prototype
In the above evaluation, participants were asked to visit sites that we sent them, and used a pre-set web history that was not populated entirely with sites that they themselves had visited. In our second evaluation, we examined the performance of CWH against the default history feature of Firefox 3 using only sites that participants had personally visited. We evaluated CWH versus Firefox 3's history by measuring the speed and accuracy in which participants were able to find web pages using them. The tradeoff here is that this approach is less controlled, with people visiting very different sites and being asked to revisit different sites as well. However, our second evaluation is also more characteristic of what people might do in real life situations.

Method
Twelve participants partook in two sessions, one hour long each and spaced a week apart. Participants were between 18 to 33 years of age. Six participants used Firefox, two used Internet Explorer, and four used both browsers equally.

In the first session, participants were asked to browse the web as they normally would for an hour. We also asked people to think out loud while browsing. Participants used our computers so that CWH could record their browsing histories. Each participant started out with his or her own browser profile and a clean history. In the first half of the second session, participants were asked to browse the web as they normally would for half an hour, with their own profiles on our computers. For the second half of the second session, participants were informed that this was a study regarding web history, at which point they were asked to revisit web pages from the past two sessions.

We selected web pages for participants to re-visit based on

if the user displayed interest in the page (e.g. laughed at the content), if the user spent a long time on the page, or if the user made comments about the page (e.g. "that's not what I want"). Each web page was described to the participant either by the participant's main purpose in visiting the page (e.g. looking for a laptop), the type or topic of the page (e.g. the Olympics), or comments the participant made about the page, rather than giving them titles or domain names.

Participants were asked to revisit five to eight web pages, depending on how many pages they visited in total across both sessions. Each participant used either Firefox 3's default web history feature or CWH, but not both (between-subjects). Each method of revisitation was evaluated by measuring four criteria: (1) the time needed to access Firefox history or CWH, (2) the time needed to find the web page, (3) the number of other web pages visited before reaching the desired page, (4) and the number of web pages
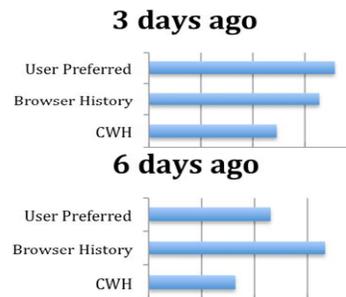

Figure 6. Y-axis represents different methods of revisitation, and the X-axis represents the number of steps for refinding and revisiting a web page during the first evaluation of the first prototype.

	Avg. time to find page	Avg. incorrect web pages visited	Avg. fraction of web pages found
Browser History	Mean: 68 ± 36 sec	Mean: 1.093 ± 1.815	Mean: 0.781 ± 0.417
Using CWH	Mean: 31 ± 13 sec	Mean: 0.4 ± 1.265	Mean: 0.7 ± 0.483

Table 3. Comparison between participants who used the browser history and the participants who used the CWH prototype to look for web pages (p=0.06)
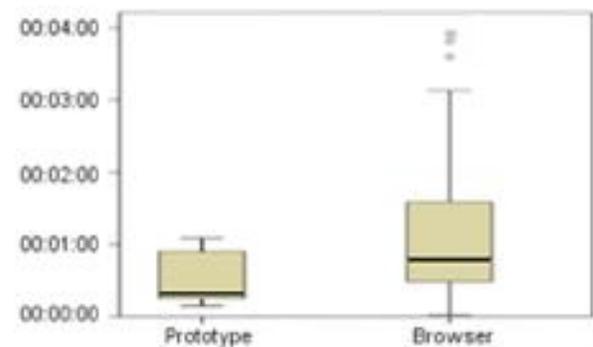

Figure 7. Time to find a web page using first prototype of CWH vs. Firefox 3's browser history (p=0.06).

that were found. The first criterion tested whether our prototype improved history's visibility and accessibility, while the other criteria evaluated the ease of revisitation.

Key Findings
Table 3 shows detailed results of how the participants who used Firefox 3's default history performed compared to participants who used Contextual Web History.

On average, participants who used Firefox's history took longer (p=0.06) to find the desired web pages and visited more incorrect web pages (Table 3, Figure 7). Many participants complained that there were too many similar entries in Firefox's history, as we saw earlier with Figure 5. Participants using Firefox's history also frequently scrolled over all entries repeatedly, or clicked on an entry that they rejected before, because they lost track of which entries they examined. Overall, CWH decreased the time and effort needed to find a web page by improving precision in displaying candidate pages and by providing faster access. While participants were able to find pages quickly, they often found it difficult to search for specific terms.

SECOND PROTOTYPE AND EVALUATION
In the second prototype, we improved the usability and overall design based on findings from our previous prototype, making CWH more visible, easy to use, and coherent with the rest of the browser. While the types of information (full text content, thumbnails, keywords, URLs, etc) remained the same, we made several changes to the front-end. First, we integrated CWH more with the browser, moving all the core features—browsing, searching, and displaying results—into a sidebar (Figure 1). To address participants' unease using the search, users can now browse the entire history and search results with thumbnails, titles, urls, and descriptions. Furthermore, from the first prototype evaluation, we found that people still tended to use their favorite search engines to revisit websites, as it was a process that they were familiar with. To accommodate this, our second prototype alerts the user to relevant history results when searches are made in web-based search engines (Google in our current prototype, see Figure 8).

CWH was implemented as a Firefox 3 extension using JavaScript, XUL, and HTML. Earlier iterations of CWH were built from scratch, while the latest prototype extends Scrapbook, a Firefox 3 extension. We chose Scrapbook because it saves copies of visited web pages and allows full-text search. The features we added to Scrapbook include storing and displaying thumbnails for each visited page, displaying CWH results whenever a user searches on a search engine (Figure 8), and having direct CWH search input in the toolbar. A thumbnail is saved by taking an automatic low-resolution screenshot of the page using the Canvas property in Mozilla API. We embedded CWH results into Google's search result page by modifying the DOM with JavaScript. Keywords are extracted from the

Figure 8. Screenshot of CWH results integrated into Google's search results. This technique greatly improved visibility and fits into people's existing process of re-finding previously visited web pages.
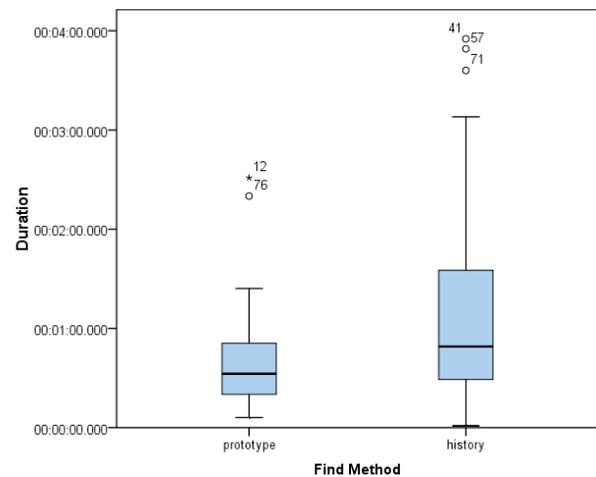

Figure 9. Average time to find a web page using CWH prototype vs. Firefox (stars and circles are outliers).

title, URL, meta tag, and content of the web page by discarding HTML markup and discarding words less than three letters.

Evaluation of the Second Prototype

Method
We met with eight participants over two separate one-hour sessions spaced a week apart. Participants were paid $30 USD after the second session. Participants were aged between 19 and 23 and four were male. Six participants used Firefox as their primary browser, one user used Internet Explorer, and one used Flock. We don't believe that higher numbers of Firefox users would bias our results because data from our and past work suggest that history is not used often in any browser. The evaluation method remained the same as for the first prototype, with our goal being to evaluate and improve the CWH user interface.

Key Findings
Six out of eight participants browsed with the prototype instead of searching with the prototype to find the sites they were looking for; browsing also had a high success rate of

finding the web pages. The setup of the study called more for the browsing approach because browsing serves to find a vague concept while searching serves to quickly find a specific page. For the first five participants, the prototype alerted the users with its results in a search engine results page in a pop-up form at the bottom right side of the browser. However, all users thought the pop-up was an advertisement. For the later three participants, we displayed our results in the search engine results page instead of displaying it as a popup (Figure 8). One participant still ignored the results because he mistook them for sponsored links, but the others quickly understood that these were history results. Some participants were misled by the ranking algorithm used in the backend. These participants expected search terms matching the title of a web page to yield higher ranking results than if the terms only matched words in the body. The algorithm we used, however did not discriminate between where the searched words appeared.

We also compared the elapsed time for finding web pages using the second CWH prototype and Firefox 3's history. Since the protocol we used for evaluating the second CWH prototype was the same as the previous study, we compared current participants, who used CWH, with the participants from the previous study that used Firefox's history.

Figure 9 shows that participants who used CWH (*M=47.2s*, *SD=40.0s*) were slightly faster than the participants who used browser history (*M=68.6s, SD=56.9s*), but only marginally significantly so, *t(44)=-1.80, p=0.079*. Our second prototype did not vastly improve speed from the first prototype, because, to our surprise, most participants preferred to browse through a listing of websites rather than search for a single website. One participant, while looking for a web page, explained that he had an idea of what to search for, but he wanted to browse through the list first since he found it easier than searching.

We also found a significant difference in the success rate of finding web pages, *t(55)=3.87, p<0.001*, between participants who used CWH (*M=100% success*, *SD=0%*) and participants who used Firefox history (*M=79% success*, *SD=41.4%*). During the user studies of the first prototype, the only way to access CWH entries was to conduct a search, and success was highly dependent on a good search query. Current prototype lets users search or browse through CWH entries, significantly improving success rate over browser history while taking less time.

CONCLUSIONS AND FUTURE WORK
Revisitation is a key part of web browsing. However, current web browser histories are not effective for revisiting web pages. From formative user studies and from past work, we found that there was a lack of awareness of the feature, and that people generally found existing histories too hard to use. We also identified cues that people could recall about the websites they visited, and experimented

with thumbnails of web pages to find a size that was effective in helping people identify sites.

Informed by these results, we iteratively developed and refined Contextual Web History over multiple paper and interactive prototypes, to understand the right set of basic features to support the process of re-finding information. We combined several relatively straightforward ideas for helping people find the history feature, browse and search through history, see relevant information about past pages, and quickly find previously visited web pages. We developed and evaluated two prototypes of CWH, and provide evidence through user studies that CWH is effective for helping people find web pages again.

One way of improving CWH is to integrate it directly into the URL locator bar, showing thumbnails and metadata about visited web pages based on user entered search terms, dates, URL, or page title snippets. Another improvement would be to allow usage of web pages that were linked to and from the target page to aid search, since past work has shown that how a person got to a page is useful. There are also several studies that could be done to further validate CWH's design. These include testing CWH on large personal histories spanning longer time periods, as well as releasing CWH and getting qualitative feedback from users as well as quantitative feedback from logs.

REFERENCES
1. BrowseBack: Visual Web History You Can Search. http://www.smileonmymac.com/browseback/
2. MindRetrieve - Search Your Personal Web. http://www.mindretrieve.net/
3. Abrams, D., R. Baecker, and M. Chignell. Information archiving with bookmarks: personal webspace construction and organization. In Proceedings of *CHI'98*: ACM Press. pp. 41-48 1998.
4. Adar, E., J. Teevan, and S. Dumais. Large Scale Analysis of Web Revisitation Patterns. In Proceedings of *Twenty-sixth annual SIGCHI conference on Human factors in computing systems (CHI 2008)*. pp. 1197-1206 2008.
5. Aula, A., N. Jhaveri, and M. Käki. Information Search and Re-access Strategies of Experienced Web Users. In Proceedings of *WWW 2005*. pp. 583-592 2005.
6. Ayers, E. and J. Stasko. Using graphic history in browsing the World Wide Web. In Proceedings of *WWW 95* 1995.
7. Blanc-Brude, T. and D.L. Scapin. What do People Recall about their Documents? Implications for Desktop Search Tools. In Proceedings of *Intelligent User Interfaces (IUI'07)*. pp. 102-110 2007.
8. Byrne, M.D., B.E. John, N.S. Wehrle, and D.C. Crow.

The tangled web we wove: A taskonomy of WWW use. In Proceedings of *Human Factors in Computing Systems: Proceedings of CHI 99*. pp. 544-551 1999.

9. Catledge, L.D. and J.E. Pitkow. Characterizing browsing strategies in the World-Wide Web. In Proceedings of *WWW 1995*. pp. 1065-1073 1995.

10. Cockburn, A. and S. Greenberg, Issues of Page Representation and Organisation in Web Browser's Revisitation Tools. *Australian Journal of Information Systems* 2000. **7**: p. 120--127.

11. Cockburn, A., S. Greenberg, B. McKenzie, M. Jasonsmith, and S. Kaasten. WebView: A Graphical Aid for Revisiting Web Pages. In Proceedings of *OZCHI'99 Australian Conference on Human Computer Interaction* 1999.

12. Cockburn, A. and B. McKenzie, What Do Web Users Do? An Empirical Analysis of Web Use. *International Journal of Human-Computer Studies* 2001. **54**(6): p. 903-922.

13. Cockburn, A., B. McKenzie, and M. Jason-Smith, Pushing Back: Evaluating a New Behaviour for the Back and Forward Buttons in Web Browsers. *International Journal of Human-Computer Studies* 2002. **57**(5): p. 397-414.

14. Czerwinski, M., M. van Dantzich, G.G. Robertson, and H. Hoffman. The Contribution of Thumbnail Image, Mouse-over Text and Spatial Location Memory to Web Page Retrieval in 3D. In Proceedings of *Interact'99*. pp. 163-170 1999.

15. Hightower, R.R., L.T. Ring, J.I. Helfman, B.B. Bederson, and J.D. Hollan. Graphical multiscale Web histories: a study of padprints. In Proceedings of *HYPERTEXT '98: Proceedings of the ninth ACM conference on Hypertext and hypermedia*. pp. 58-65 1998.

16. Jones, W., H. Bruce, and S. Dumais. Keeping found things found on the web. In Proceedings of *The tenth international conference on Information and knowledge management (CIKM01)*. pp. 119 - 126 2001.

17. Jones, W., S. Dumais, and H. Bruce, Once Found, What Then? A Study of "Keeping" Behaviors in Personal Use of Web Information. *ASIST 2002* 2002. **39**(1): p. 391-402.

18. Jones, W.P. and S.T. Dumais, The Spatial Metaphor for User Interfaces: Experimental Tests of Reference by Location versus Name. *ACM Transactions on Office Information Systems* 1986. **4**(1): p. 42-63.

19. Kaasten, S. and S. Greenberg. Integrating back, history and bookmarks in web browsers. In Proceedings of *CHI '01: CHI '01 extended abstracts on Human factors in computing systems*. Seattle, Washington. pp. 379-380 2001.

20. Lansdale, M.W., Remembering about Documents: Memory for Appearance, Format, and Location. *Ergonomics* 1991. **34**(8): p. 1161-1178.

21. LeeTiernan, S., S. Farnham, and L. Cheng. Two methods for auto-organizing personal web history. In Proceedings of *CHI '03: CHI '03 extended abstracts on Human factors in computing systems*. Ft. Lauderdale, Florida, USA. pp. 814-815 2003.

22. Milic-Frayling, N., R. Jones, K. Rodden, G. Smyth, A. Blackwell, and R. Sommerer. Smartback: supporting users in back navigation. In Proceedings of *WWW'04*: ACM Press. pp. 63-71 2004.

23. Obendorf, H., H. Weinreich, E. Herder, and M. Mayer. Web Page Revisitation Revisited: Implications of a Long-term Click-stream Study of Browser Usage. In Proceedings of *CHI 2007*. pp. 597-606 2007.

24. Robertson, G., M. Czerwinski, K. Larson, D.C. Robbins, D. Thiel, and M.v. Dantzich. Data mountain: using spatial memory for document management. In Proceedings of *11th annual ACM symposium on User interface software and technology (UIST 1998)*. pp. 153-162 1998.

25. Suh, B., A. Woodruff, R. Rosenholtz, and A. Glass. Popout Prism: Adding Perceptual Principles to Overview+Detail Document Interfaces. In Proceedings of *The SIGCHI conference on Human factors in computing systems (CHI2002)*. pp. 251-258 2002.

26. Tauscher, L. and S. Greenberg, How People Revisit Web Pages: Empirical Findings and Implications for the Design of History Systems. *International Journal of Human Computer Studies* 1997. **47**(1): p. 97-138.

27. Teevan, J. The re:search engine: simultaneous support for finding and re-finding. In Proceedings of *UIST '07: Proceedings of the 20th annual ACM symposium on User interface software and technology*. Newport, Rhode Island, USA. pp. 23-32 2007.

28. Teevan, J., E. Adar, R. Jones, and M. Potts. Information re-retrieval: Repeat queries in Yahoo's logs. In Proceedings of *SIGIR '07* 2007.

29. Weinreich, H., H. Obendorf, H. Herder, and M. Mayer. Off the Beaten Tracks: Exploring Three Aspects of Web Navigation. In Proceedings of *WWW 2006*: ACM Press. pp. 133-142 2006.

30. Wexelblat, A. and P. Maes. Footprints: history-rich tools for information foraging. In Proceedings of *Seventeenth annual SIGCHI conference on Human factors in computing systems (CHI 99)*. Pittsburgh, Pennsylvania, United States. pp. 270-277 1999.

31. Woodruff, A., A. Faulring, R. Rosenholtz, J. Morrsion, and P. Pirolli. Using thumbnails to search the Web. In Proceedings of *Nineteenth annual SIGCHI conference on Human factors in computing systems (CHI 2001)*. Seattle, Washington, United States. pp. 198-205 2001.