Personal Web Library: Organizing and Visualizing Web Browsing History

Abstract

Purpose - Web browsing has become an indispensable activity for many people to support their daily work, research, and entertainment. Modern web browsers all provide a history function that allows users to see a list of URLs they have visited in chronological order. The history log contains rich information but is seldom used because of the tedious nature of scrolling through long lists. This paper proposes a new way to improve users' web browsing experience by analyzing, clustering, and visualizing their browsing history.

Design/methodology/approach - We developed a system called Personal Web Library to help users: 1) develop awareness of and understand their web browsing patterns, 2) identify their topics of interest, and 3) retrieve previously visited web pages more easily. We describe the design process and the implementation of a high-fidelity functional prototype.

Findings - Results from a user study indicate that users appreciate being able to see the visualization of their browsing history, and suggest that users tend to be interested in improving their time management skills based on an enhanced awareness of their own browsing habits.

Originality/value - Providing users with interactive visualizations of their browsing history can facilitate personal information management, time management, and other meta-level activities. We hope the work we conducted can spur more research contributions in this underdeveloped yet important area.

Keywords: web browsing history; personal informatics; information visualization; topic clustering; user experience

Paper type: Research Paper

---

1. Introduction

As more and more activities rely on the World Wide Web, people search, browse, and read many web pages daily in both work and personal contexts. Most modern web browsers have a history function that records and displays previously visited pages. In a short period of time (e.g., one or two weeks), a user may visit hundreds to thousands of web pages, accumulating a long list in their browser's history log. This browsing history has value, as it may reflect the user's interests, needs, and daily habits (Van Kleek, Moore, Xu, & Karger, 2010a), the understanding of which is an important part of personal informatics. However, the potential value in this data is often left untapped. Won et al (Won, Jin, & Hong, 2009a) note 2 reasons why browsing history is rarely used: (1) the history is hidden in the browser, and (2) current implementations of the history log do not adequately support users in engaging with their history. Typically, web browsers present only a list-based, textual representation of the history log, which comprises a long list of URLs ordered by visiting time. These long lists generally do not help users identify or extract useful information and are cognitively demanding. One possible solution to this problem is to visualize browsing history, which can help overcome limitations inherent in list-based textual representations (Larkin & Simon, 1987).

In the past, researchers have explored different alternatives for visualizing browsing history. Most previous work has focused on individual web pages or on the chains that connect visited web pages. For example, Eyebrowse (Van Kleek et al., 2010a) and HistoryLane (Chtivelband, 2012) focus

on only the most visited websites. CZweb (Fisher et al., 1997; Milic-Frayling, Sommerer, & Rodden, 2003) and SessionGraphs (Mayer, 2009) use node-link graphs to show access sequences and connections among web pages. These approaches provide users with a detailed understanding of one section of the visited history, but do not provide enough of an overview to help users see the "whole picture" and understand patterns and trends over longer periods of time.

When users engage in a browsing session there is usually a theme—for example, seeking literature on a research topic, finding gifts for a holiday, or comparing products for an intended purchase. Users generally do not visit web pages randomly, but rather tend to focus on a collection of web pages around one topic in a short period of time. Although in the end users may find only a few pages containing useful information, other pages visited during the browsing session may provide a useful context for the topic and target pages. We believe that the information about the themes, URLs, and temporal patterns of web browsing, can provide the users an in-depth self-awareness toward their web browsing behaviors and habits.

In this paper, we present a comprehensive approach for organizing and visualizing web browsing history. We describe our design process, from identifying requirements through design ideation, iterating on early prototypes, and settling on a final prototype that we call Personal Web Library (PWL). We utilize a text clustering technique, and visualize browsing history information on two levels (topics and URLs) and at two time-granularities (daily in two-week view and hourly in single-day view). We present our final prototype and discuss its features. We report the results of a user study aimed at assessing the usability and overall user experience of our system. Results indicate that users appreciate being able to see the visualization of their browsing history, and suggest that users tend to be interested in improving their time management skills based on an enhanced awareness of their own browsing habits. We also found that two main interface levels worked well for users—namely, a two-week view and a single-day view.

The rest of the paper is organized as follows. In Section 2 we discuss related work; in Section 3 we describe data extraction and processing; in Section 4 we discuss our design process and outline our system implementation; in Section 5 we describe our usability evaluation, key takeaways, and future work; finally, in Section 6 we present conclusions.

2. Related Work

2.1. Web Browsing History

In almost all modern web browsers, including Chrome, Firefox, and Edge, the URL of each visited web page is recorded in a history log. This history log can be very useful and has been the subject of investigation and discussion for decades. Since the mid-1990s, there have been many different approaches to analyzing, using, and representing browsing history. For instance, some work looked at analyzing personal information in bookmarks. Other research tried to analyze why and how users create, use, and organize bookmarks, and provided recommendations for improvement (Abrams, Baecker, & Chignell, 1998). Some studies focused on exploring users' revisitation patterns and web browsing behaviors based on log files (Adar, Teevan, & Dumais, 2008; Tauscher, 1996). More recently, some work has been aimed at building personalized search engines using browsing history to provide more accurate personalized search results (Kumar & Sharan, 2014; Matthijs & Radlinski, 2011).

Combining browsing history with human behavior analysis and visual representation methods to design better revisitation tools is another area of research focus. Representations such as graphical tree structures (Ayers & Stasko, 1995), basic graphs (Wexelblat & Maes, 1999a), 3D tools (Cugini & Scholtz, 1999), and the combination of more complex visualizations (Cernea, Truderung, Kerren, & Ebert, 2013a) have been explored. Some research has also examined the use of graphic tools to share history information with others for social purposes (Van Kleek, Moore, Xu, & Karger, 2010b).

Modern browsers provide a search function to help users retrieve visited URLs from the history log. However, one common problem is that it is difficult for users to remember keywords previously used (Won, Jin, & Hong, 2009b). Some attempts have been made to provide cues to users—for example, using thumbnails of visited pages as contextual cues to expedite revisitation (Gandhi, Kumar, Bederson, & Shneiderman, 2000; Won et al., 2009a) . Won et al. (Won et al., 2009b)

generated the following ranking of website features that could easily be remembered by users (from easiest to most difficult): colors and images, visual structure and layout, time spent on the site, logos, animated content, page title, the URL's domain name, and the URL's path.

2.2. Approaches to Visualizing Browsing History

Information visualization plays a significant role in the extant research on web browsing history. Although web browsing data are abstract, visual representations of the data can reinforce human cognition and help users develop a better understanding of their browsing history and habits (Mayer, 2009). The application of information visualization techniques to browsing history data has become more popular in recent years. Examples of previous work on visualizing browsing history are characterized below according to their visualization approach:

Node-link visualizations: In VISVIP (Cugini & Scholtz, 1999), researchers used website nicknames as nodes and linked them according to the network relationship. This tool mapped the 3D visiting paths of web pages onto 2D visualizations. This strategy not only clearly presents the websites' information structure, but also provides an interactive way for users to select a subject's path for viewing. This is an improvement from studies of other similar systems, such as Footprints (Wexelblat & Maes, 1999b) , which showed paths of several users at a time.

Spatial visualizations: TRAILS (Yu & Ingalls, 2011) is a more complex visualization that combines browsing frequency, favicons, time, connections, tags defined by users, and sorting functions. It uses creative visual cues to show users their browsing frequency, connections among visited websites, and category of the browsing history with intuitive, unobtrusive representations, and minimal cognitive load. The multiple visualization views provide users with a better overview and a more interesting experience, according to the conducted user study. However, efficiency could be an issue in this approach, as some users indicated they had no time to manually tag the rings in the visualization and play with them.

Temporal visualizations: WebComets (Cernea, Truderung, Kerren, & Ebert, 2013b) is a visualization that focuses on supporting the understanding and analysis of parallel browsing behavior based on navigation among multiple tabs. Time lines, tails, pie charts, links, and color coding are used to represent active time, focus time, and connections among pages. WebComets is a comprehensive visual analytics system for experienced visual analysts. As a result, it is difficult for novice users to understand the meaning of the visual elements. Eyebrowse (Van Kleek et al., 2010a) tracks and visualizes the most visited webpages using different types of representations to communicate trends over time. Eyebrowse uses a stacked bar chart to indicate the most visited URLs for different days of a week or different hours of a day, and a line graph to show the number of pages visited over the course of one week.

These examples highlight different visualization approaches that have been used to help users make sense of their browsing history. Although some work has been done, gaps still exist, especially for helping general users understand their past browsing habits. The aim of our design is to provide a tool that helps users understand these habits using interactive visualizations that incorporate keyword clustering techniques.

2.3 Keyword Clustering and Topic Detection

Multiple research efforts have been aimed at automatically assigning text documents to a set of given categories (Li & Yamanishi, 2003; Wartena & Brussee, 2008; Zu Eissen & Stein, 2002). For example, Li and Yamanishi (2003) have proposed an approach to topic analysis using a finite mixture model to represent word distribution. Latent Semantic Analysis (LSA) (Dennis, Landauer, Kintsch, & Quesada, 2003) is a common approach that extracts and represents the contextual-usage meaning of words by statistical computation. Extending LSA, Latent Dirichlet Allocation (LDA) assumes that topic distribution has a sparse Dirichlet prior, which results in a better disambiguation of words and more precise placements of documents to topics (Blei, Ng, & Jordan, 2003). Wartena and Brussee (Wartena & Brussee, 2008) utilized a k-means clustering algorithm to group keywords into topics without prior knowledge. This k-means approach is related to LSA in that both use occurrence and co-occurrence data of keywords as input. In LSA, observed keywords are decomposed into

several latent distributions using their conditional probability distributions. Words with strong mutual co-occurrence tend to have the same main latent components. In our work (and Wartena's work (Wartena & Brussee, 2008), clustering is achieved by comparing and grouping distributions of co-occurring terms. The center of a cluster is the average distribution of the co-occurrence distributions, which is comparable to a latent component in LSA (Wartena & Brussee, 2008).

Within the past decade, researchers have developed visual analytics tools to aid in text analysis. For example, TIARA (Wei et al., 2010) employs topic analysis techniques to summarize topics into a sets, then uses several interactive text visualization techniques to show these topics. TextFlow (Cui et al., 2011) uses topic mining techniques to detect various evolution patterns that emerge from multiple topics. LeadLine (Dou, Wang, Skau, Ribarsky, & Zhou, 2012) uses topic modeling to identify meaningful events in news and social media text with the help of stream-like visualizations. TimeLineCurator (Fulda, Brehmer, & Munzner, 2016) extracts events from text and visualizes them on a timeline. Researchers have also proposed various interactive visualization techniques to aid in different aspects of text analysis, such as Parallel Tag Clouds (Collins, Viegas, & Wattenberg, 2009) and semantic interaction (Endert, Fiaux, & North, 2012). Although visual analytics tools and techniques have been developed in recent years to address numerous challenges posed by text analysis, they have not been aimed specifically at addressing issues related to web browsing history. Our work aims to contribute to this gap in the literature.

3. Data Extraction and Processing

All modern web browsers provide a history function that records and stores browsing history. In this project, we used the browsing history file created by Google Chrome. The file is in SQLite format, contains a timestamp, URL, and website titles. To get more information on the page, we wrote as script to retrieve keywords from the title, keywords, and description metatags of each URL by revisiting the webpage. Stop words (e.g., "the," "is," "a") were removed. We used this simple approach for the following reasons: 1) although many web pages contain large chunks of meaningful semantic text, a major portion of web pages do not strictly have the information. They may just be a simple collection of resources such as product pages in e-commerce sites, music, or videos. Such content may not suitable for LSA or LDA. 2) In most cases these meta-tags are manually entered by website administrators, or automatically generated by the website, to reflect the page content. Although current search engines do not use this information for ranking, they are still important for search engines to index and retrieve them. And 3) identifying a set of keywords from a webpage, or determining the weight of each keyword on each document, have already been important research topics by themselves (Matsuo & Ishizuka, 2004). Even search Engines like Google and Bing will come up with different webpages using the same search term. The focus of our work is finding the most representative keywords for webpages. We consider these keywords to be a rough description of the topics or themes of a webpage.

From each webpage, it is possible to obtain several to tens of keywords. From two weeks' worth of browsing history from a single user, we were able to extract several hundreds of keywords. As shown in Figure 1, one URL may contain many keywords, and one keyword maybe used by many URLs. It is possible to group either URL or keywords using these co-occurrence relations. Here we chose to cluster keywords since they are commonly used when searching, and may be useful in indicating users' main topics.

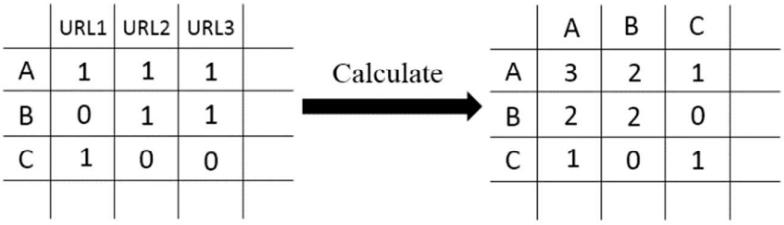


Figure 1. Closeness of keywords. The closeness of keywords A and B is two since there are two URLs (URL2 and URL3) containing both keywords.

To make the number of keywords more manageable, we employed a K-means clustering algorithm to group them into themes. If two keywords were used on the same webpage, these two keywords were considered to be related. The more webpages containing multiple keywords, the "closer" these two keywords were. The "closeness" is determined by the number of URLs containing both keywords (see Figure 1). With this strategy, we could group related keywords together into clusters. The K-means algorithm requires a pre-defined number of groups to start clustering. Using the elbow method (Kodinariya & Makwana, 2013), and testing with several users' browsing history, we attempted to cluster the keywords from 10 groups to 50 groups. We found that the most effective number of groups was approximately 20. With this number, most groups are relatively well defined with clear boundaries—except one group that contains keywords not fitting into the other clusters. No matter how many groups are set (e.g. 50 groups), there is always such a group. We examined the group and noticed it contains many isolated keywords that do not appear together frequently in web pages. According to our observations, we noticed that users sometimes do browse the web randomly without a clear focus, which explains this 'leftover' group.

During development we also tried the agglomerative hierarchical clustering algorithm. We manually examined the result and found that at higher clustering levels, most clustering results were hard to explain, as many unrelated keywords were grouped together. Our testing showed less meaningful results than with K-means clustering. Steinbach et.al (Steinbach, Karypis, Kumar, & others, 2000) compared several popular clustering algorithms, including "standard" K-means, "bisecting" K-means, and agglomerative hierarchical clustering on text clustering. Both K-means methods perform better than hierarchical method. K-means methods also have a better time complexity of O(n). Bisecting K-means performs slightly better than standard k-means on most cases with several exceptions. Although future work may tweak the clustering method for improved performance, the current aim is to build an interactive visualization interface to help users understand their browsing history. Thus, since our focus is not on clustering algorithms per se, we simply adopt the most straight forward K-means method.

Our next step was to associate web pages with keyword groups. Sometimes a webpage contains keywords belonging to multiple groups. In this situation, the page belongs only to the group containing the most keywords from the page. Subsequently, each keyword cluster was associated with many web pages and corresponding visiting timestamps. We then designed a visualization system to help users gain insight into and reuse their browsing history by viewing the time and topics of web visits at two main levels of temporal granularity.

4. System Design and Implementation

Based on the above noted scarcity of research, we set out to devise strategies for visualizing users' web browsing history and implement them in an interactive visualization tool. In the following subsections, we describe our design process, including setting initial goals based on user tasks, design ideation, concrete design iterations, representation and interaction design, and implementation details.

4.1 Goals and User Tasks

After understanding the specific domain situation (e.g., web browsing history), the next typical step in the visualization design process is to understand the tasks that a visualization tool should support (Munzner, 2009). Although there is no consensus on the definition of a task in the visualization literature, tasks are widely recognized as central to visualization design and use (Rind, Aigner, Wagner, Miksch, & Lammarsch, 2016). Designers can identify relevant tasks and can design interactions that help users achieve those tasks. Based on our analysis, we determined that our system should support users to:

- identify prominent topics in their web browsing history
- view their history at different levels of temporal granularity
- identify patterns and trends within topics
- locate and revisit particular web pages
- choose and apply different approaches to narrowing the scope of URL lists

In addition to the tasks that we aimed to support, we also aimed provide a pleasant experience to the user.

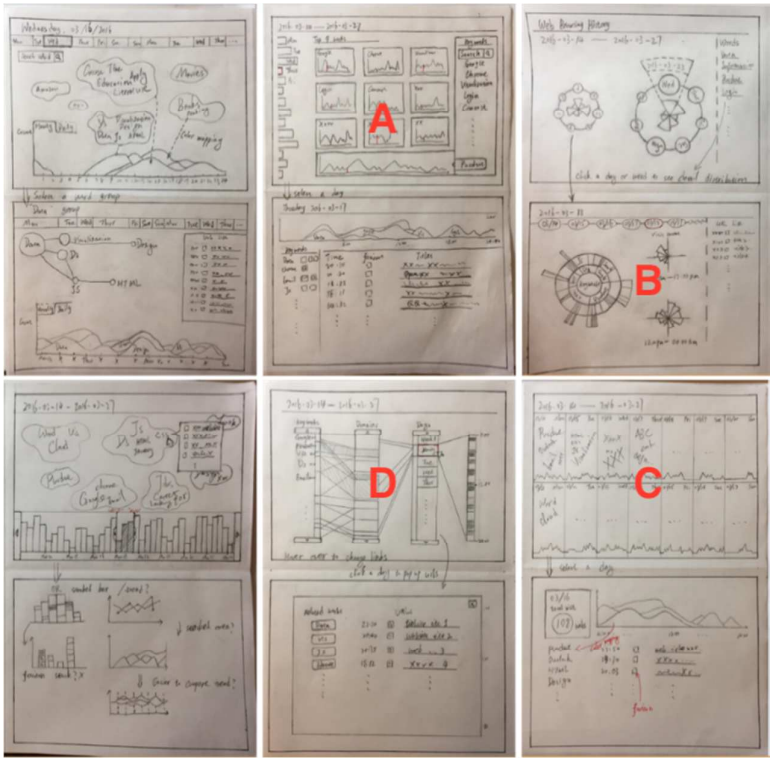


Figure 2. Sketches of design ideas during the ideation phase

4.2 Design Ideation

Ideation, which refers to the creative process of generating and communicating ideas, is an important component of any design process. One of the most widely used methods for ideation is sketching. Sketching supports design ideation by allowing ideas to be represented at different levels of abstraction, and supports reinterpretation and design discussion (Purcell & Gero, 1998).

Based on characteristics of the available data, and on the goals and tasks described above, we first engaged in brainstorming sessions to generate multiple ideas. We then aggregated the ideas and created various interface sketches (see Figure 2). We explored the use of different visualization techniques, focusing on their affordances and limitations. For example, the small multiples technique (Figure 2 A) allows users to see trends among the most-visited topics, but the user has to mentally combine all temporal curves to get an idea of overall time usage of all cells; the sunburst technique (Figure 2 B) shows the distribution of words and their hierarchical relationships, but it is hard to accommodate many words in the graph directly so that the user has to use interaction (e.g. click) to see the detailed words in each segment; the calendar representation (Figure 2 C) provides daily trends in a familiar format but each cell will be too small to deliver sufficient details; the Sankey diagram (Figure 2 D) may show topic transitions while the user browses a website, but from our user study, people often switch topics randomly. Thus, transitions seem to carry little meaningful information. We explored different options for representing time, including various circular and linear forms. We also considered relationships among time, visit count, keywords, and websites being represented with multiple linked views and coordinated interactions across them. Another important factor in our choice of visualizations is their ease of use. Our target users are regular browser users who do not necessarily have information visualization experience. Thus, we also wanted to choose simple intuitive forms that are easy to understand.

4.3 Design Challenges and Conceptual Solutions

While sketching and discussing these initial designs, we realized that there were two major design challenges. The first challenge was how to provide a useful overview of keywords of different frequencies and time distributions, considering the potentially large number of keywords to visualize—a number that varies significantly depending on the users. After comparing different visualization methods and their affordances, we chose to cluster the keywords and represent them with wordClouds. An experiment by Rivadeneira et al. (2007) suggests that wordClouds have the following benefits: (a) helping users locate or determine the existence (or lack thereof) of a concept; (b) supporting browsing, often with no specific target in mind; (c) helping users form an impression or get the "gist" of the content, especially by drawing awareness to very prominent or not so prominent topics; and (d) recognizing which of several possible sets of information is being represented, and matching the information to a personally meaningful event in the past. Due to these benefits, and the lack of such benefits in many other visualization techniques, we decided to make use of wordClouds. All four benefits are in alignment with our goals for the Personal Web Library. In our wordCloud, each keyword's font size encodes the frequency with which the keyword was visited by the user. More frequently visited keywords stand out due to their large size, helping users quickly get the "gist" of the content.

The second challenge we identified was how to effectively represent the visitation trends and patterns of many groups. Being aware of the trend of each topic group is more meaningful than seeing only the trend of a single keyword. Although not always the case, for our purposes a circular representation is not as effective as linear representation in communicating temporal trends. As can be seen in the circular design in Figure 2, although the daily difference could be shown, the trend of each topic group still could not be perceived effectively. Furthermore, we realized that finding the starting point of a trend in a circular representation is not as easy as in a linear format. We then considered the suitability of the following representational strategies: small multiples, multi-series line chart, stacked bar chart, and ThemeRiver, each of which has inherent tradeoffs. For example, small multiples can show trends of different series, but the number of series is still limited by the screen space. Stacked bar graphs not only show trends of different groups like multi-series line charts, but also allow users to view the total visiting trends of all groups. Although stacked bar charts can encode individual data points more precisely and with a finer time granularity, the ThemeRiver, also known as streamgraph, integrates some basic patterns of mapping data to visual forms that have affordances which other techniques do not (based on Sedig and Parsons's recent pattern-based framework (Sedig & Parsons, 2016)):

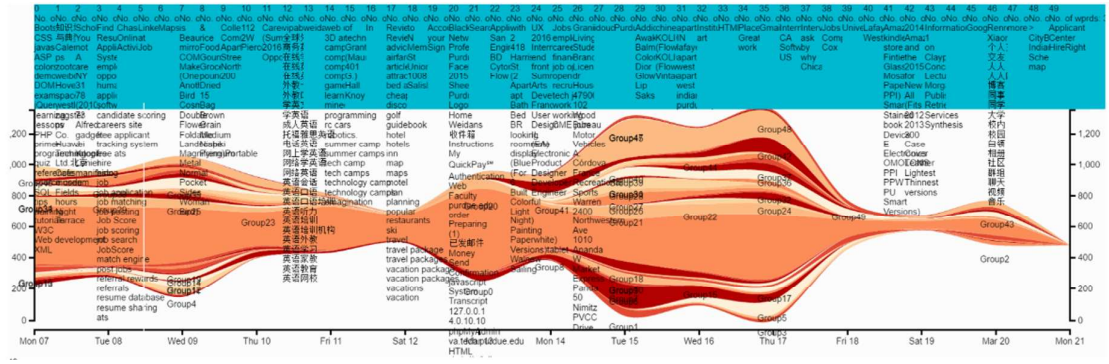
- fusing individual data points into a unified visual form, which can provide a sense of wholeness and continuity of the data
- stacking the "currents" in the river allows co-concurrent data to be viewed easily
- embedding the representation within a coordinate system provides an external frame of reference that gives meaning to spatial locations

Because of the above characteristics of the ThemeRiver, users tend to find it easier than alternative techniques to follow trends, especially macro ones (Havre, Hetzler, & Nowell, 2000). As a result, we decided to integrate the wordCloud and ThemeRiver techniques.

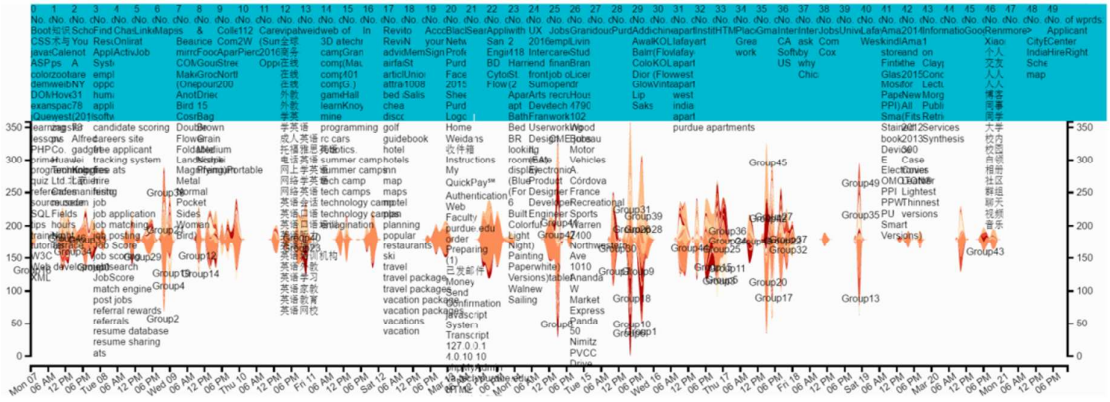
4.4. Early Design Attempts

After deciding on appropriate visualization techniques, we created early prototypes of the interface using real datasets of browsing history. Trends for each topic group were shown in the ThemeRiver using different colors. When developing the ThemeRiver, we decided to place it in the center of the interface, because the vertical proximity of the sub-rivers can make it easy for users to judge relative widths (Havre et al., 2000). Havre et al. note that users can easily perceive flow patterns and changes because of the similar symmetry around the horizontal axis of the graph (Havre et al., 2000). The initial ThemeRiver provides an overview of browsing history, which is often a useful first step in visualization-based information seeking (Shneiderman, 1996). Figure 3 depicts some of our early design attempts at integrating the wordCloud and ThemeRiver. The top and middle prototypes show our explorations with the ThemeRiver and plain keywords without the wordCloud. The top

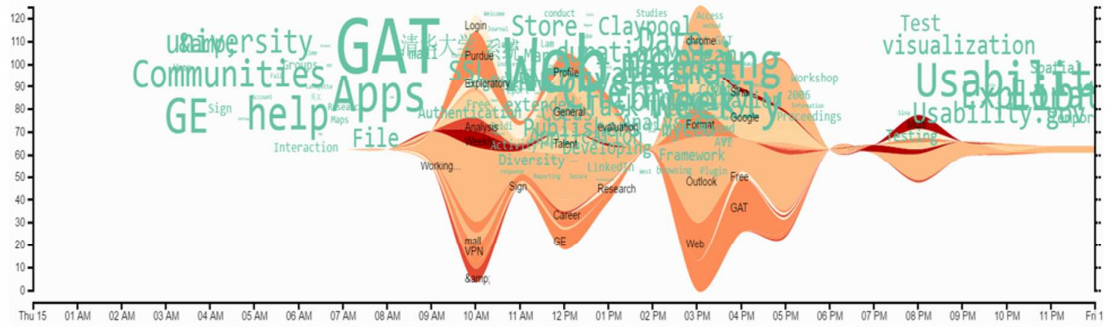
prototype shows visits aggregated by day; the middle prototype shows visits aggregated by the hour; the bottom prototype shows our experimentation with a different layout where the wordCloud was superimposed on the ThemeRiver.



(a)



(b)



(c)

Figure 3. Early design attempts at integrating wordCloud and ThemeRiver: visits aggregated by day (a); aggregated by hour (b); a different layout where the wordCloud is superimposed on the ThemeRiver (c).

4.5. Visualization and Interaction Design

After many design iterations with informal testing, we arrived at a stable prototype. Figure 4 is a screenshot of the main interface of our prototype, which shows the 2-week history of a user's browsing. The 2-week view comprises four sections: 1) a wordCloud to show the most popular keywords; 2) a ThemeRiver to show temporal trends of keyword groups; 3) a filled-area plot to show web visits on particular keyword groups; and 4) a list of URLs that are related to a chosen keyword.

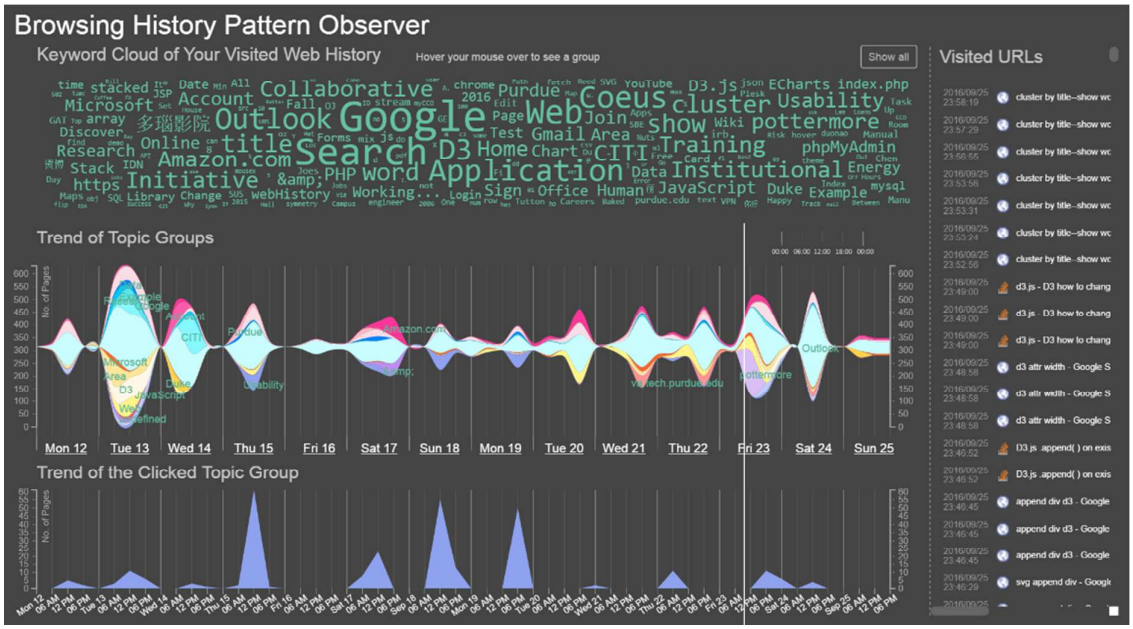


Figure 4. Stable prototype of the Personal Web Library interface showing a two-week view

In the ThemeRiver, different keyword groups are visualized as "currents" with different colors. The width of a current represents the number of visits at a given time. While designing the ThemeRiver, we had to pick a time interval to aggregate the number of webpage visits. After testing 1 day, 1 hour, and 6 hours, we found that 6 hours seems to be the best interval for the 2-week period. The 6-hour granularity naturally divides a day into morning, afternoon, evening, and night. These chunks are easily mapped to typical daily activities such as work, leisure, and sleep. From our previous user interviews and reviews of the existing literature, we knew that users find it difficult to recall exact times at which webpages were visited several days in the past. Users are more likely to recall the time at which they visited a web page in broad chunks of time—e.g., morning, afternoon, or night—with an accuracy of the early or later portion of that chunk of time (e.g. late morning).

TIARA (Wei et al., 2010), a visual analytics tool discussed previously, embeds words visually inside the currents. Although it is easier to make visual comparisons, doing so is not always possible. We found in our context of web visits, the currents are not stable in width, making it impossible to embed the words within them. In TIARA, the currents appear to be fairly wide, making it possible to embed the words within the currents. We decided on a hybrid approach, where only the first several important keywords are displayed within the currents, and the rest are displayed in the wordCloud.

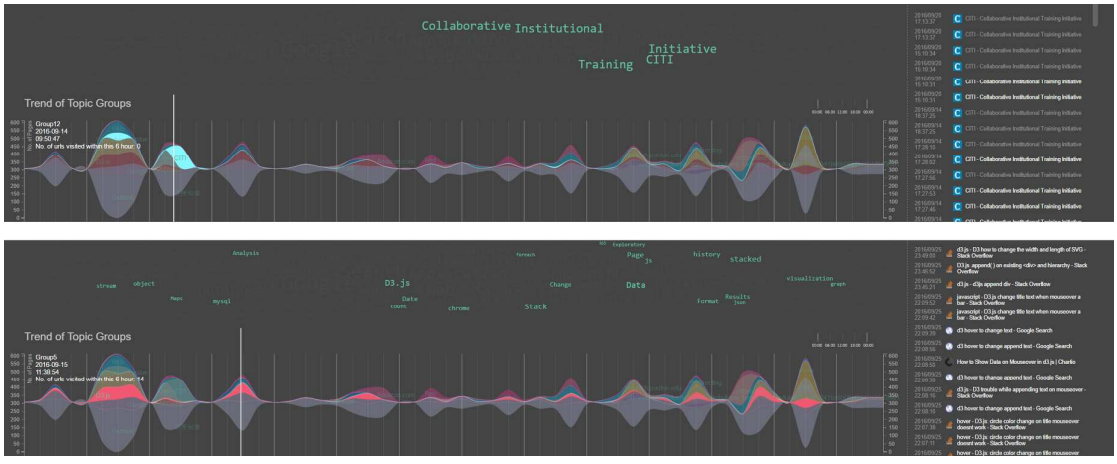


Figure 5. Two examples of keyword groups. The cursor is over one current or a keyword to highlight all keywords in the group. Visited URLs are listed on the right of the interface.

When a user moves the cursor over any keyword in the wordCloud, all keywords in the same group will be highlighted while all others are dimmed (Figure 5). Simultaneously, in the ThemeRiver, other currents will be darkened to emphasize the current keyword group. Alternatively, the user can hover over a current in the ThemeRiver, causing other currents to be darkened and only the selected group of keywords to be shown in the wordCloud. All views in the interface are dynamically linked and coordinated through interaction, which helps to overcome inherent limitations of individual views (Keim, 2002; Roberts, 2007). When a user interacts with one of the views, all others are updated accordingly. The filled-area plot at the bottom shows all visited websites related to current keywords in a smaller time granularity (every 3 hours). On the right is a list of most visited URLs that are related to the current keyword group. The URLs are ranked based on their distance from all the keywords in the group. URLs containing more keywords are ranked higher than URLs with fewer keywords. Also, more frequently visited URLs are ranked higher than less visited ones. Clicking on any URL will open the website in the browser.

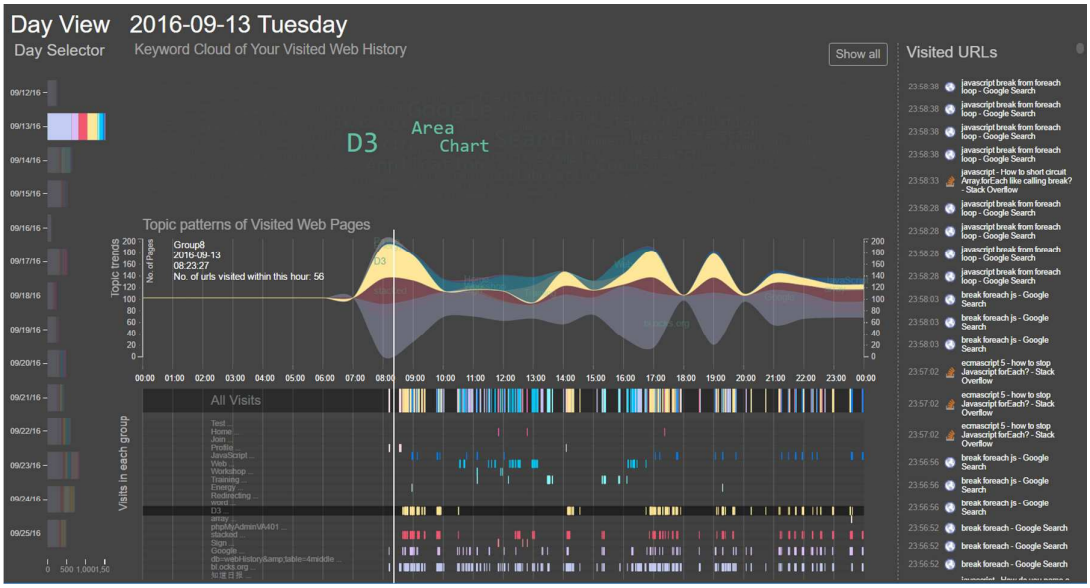


Figure 6: Personal Web Library: One-day View. The cursor is over one "current" in the ThemeRiver, causing certain keyword groups to be emphasized in the wordCloud.

In addition to the views in the main interface, users can drill deeper to see a single-day view by clicking on a day (see Figure 6). In this interface, most of the views are the same as in the two-week view—i.e., wordClouds, ThemeRiver, and the URL list. However, instead of a filled-area plot showing temporal patterns of a single keyword group, we created a detailed timeline view to show each individual web visit. The wordClouds, ThemeRiver, and URL list are the same format as the two-week view to maintain internal consistency in the interface and leverage existing mental models of users. Here we changed the granularity of the ThemeRiver to show a more detailed temporal pattern of web visits in each hour. The bottom timeline view lists the detailed time of URLs in each group. From left to right, all URL visits in the 24 hour-period are visualized using short, vertical line strokes. The color of each stroke is mapped to its attendant color in the ThemeRiver. For example, in Figure 6 the yellow color represents one group that has "D3" as its most frequent keyword. Individual URL visits are colored yellow in the timeline view, and the corresponding current in the ThemeRiver is also colored yellow. Hovering on a stroke will highlight the corresponding URL in the URL list on the right side of the interface, the relevant keywords in the wordCloud, and the current in the ThemeRiver. Clicking the stroke will open the URL in the browser.

To allow users to switch between different days quickly, we added a stacked bar chart on the left side that encodes the number of web visits per day throughout the past two weeks. This view allows users to quickly determine the relative number of visits per day and to quickly drill into any one of the days. As with the other views in the interface, all except the current day are "dimmed" to provide easy identification of the current state of the interface.

4.6. Implementation

We implemented our prototype as a standard web application using HTML5 and JavaScript on the front-end, and PHP and MySQL on the back end. Users' SQLite history logs were imported into a database located on a webserver. The server processes the log, retrieves keywords, and computes the clusters. Users can access this visualization as a regular web page through a web browser. Through username/password protection, a user can see only her own navigation history.

5. Usability Evaluation

We evaluated the system from a user experience perspective to understand how users interact with the system and develop awareness of their general browsing interests and topics. The design goal of this system is not to enhance the efficiency and accuracy of visited webpage retrieval, but to help users explore, learn, and benefit from their own browsing history. If users are looking for a specific page, and can remember enough information about it, searching with exact terms to retrieve the site is quicker than using our visualization interface. However, users often cannot remember enough detail to retrieve a page via a search query. One well-established interface design principle is that recognition is better than recall—that is, users can recognize information much more readily than they can recall it from memory. Our system, Personal Web Library, serves as a type of "recognition platform", providing different dimensions of clues for the user to recognize these browsed webpages and thus recall their browsing history.

In our evaluation, we were primarily interested in qualitative aspects of system use that are not captured in purely quantitative approaches. Quantitative approaches are useful when the aim is to measure performance, but not as useful when measuring other aspects of system use. In recent years, qualitative usability testing methods have been widely used to evaluate information visualization and visual analytics systems (Haggstrom et al., 2011; Lam, Bertini, Isenberg, Plaisant, & Carpendale, 2012; Scholtz, Plaisant, Whiting, & Grinstein, 2014). Another reason for focusing mainly on qualitative issues is that most of the tasks in which we are interested are not feasible using the typical browser history log. Of our five tasks listed in Section 4.1, only one of them can be performed easily with a traditional browser: locate and revisit particular web pages. Others, such as viewing history at different levels of temporal granularity, or identifying patterns and trends within topics, are very difficult to do with traditional history logs. As a result, it is hardly useful to measure performance on such tasks quantitatively, as there is no meaningful baseline that can be established with traditional browsers. Furthermore, we are already quite certain that traditional browsers adequately support looking up specific pages that can be recalled from memory, as their indexing and query functions are quite robust. As our system is not intended to replace such functionality, it is not meaningful to measure and compare performance values for this task. Because of these conditions, we decided that qualitative feedback would be much more useful in assessing our system.

5.1. Evaluation Settings

A usability study was conducted to evaluate whether and how our visualization prototype could help users understand their browsing history. Six STEM students were recruited for the study. Three were undergraduate students, and three were graduate students. Participants were asked to view a simple introduction to some visualizations, perform tasks using both our visualization tool and Chrome's history page, and answer a set of questions. In the evaluation, we used the participants' own browsing history to create authenticity. We imported the participants' SQLite history file into our system and processed the data. Each participant interacted with her own data. After the study, participants' history was deleted to protect their privacy. Each session lasted approximately 90

minutes. After an initial period of free, unguided exploration, in which participants could become comfortable with the system, we asked participants to complete 5 tasks. The 5 tasks are listed briefly in Table 1, and are elaborated on below.

Table 1. Brief description of the 5 tasks given to participants.

Task	Interface Focus	Description
1	2-week view	Identify important topics
2	2-week view	Identify important topics with temporal patterns
3	2-week view	Identify keywords and URLs within one topic group
4	1-day view	Identify important topics
5	Any	Retrieve a previously visited page and describe the context

The intention of Task 1 is to see what features, topics, and patterns participants could recognize in the interface. Before using the system, we asked the participants to recall the main topics in which they had been interested throughout the previous two weeks. We then allowed participants to interact with our system to recognize and recall additional topics. We recorded the number of topics each participant could recall before and after using the system. For Task 2, participants were asked to recognize several topics of interest and describe their temporal patterns. Before using the system, we asked participants to estimate how much time they spent browsing various topics on the web. Then, while using the system, we asked participants to do the same. For Task 3, we picked one group and asked participants to describe when and what web pages were visited in the group. With the presence of keywords in the wordCloud, temporal information from the ThemeRiver, and a detailed list of webpages, we asked participants to recall visited web pages, describe why they visited these pages, what they learned from these pages, and what the context of their browsing was. Task 4 was similar to Task 1, except it used the single-day view. Without the system, we asked the participants to recall topics and visiting patterns for single days. Then with the help of the system, participants checked each day again. For Task 5, we picked one visited URL from participants' history, provided some simple clues about the page's content, and asked participants to retrieve the page without using keyword search. We gave participants 2 minutes to find the page.

We used a think-aloud protocol (Ericsson & Simon, 1998) and a post-task interview during the study. Participants' processes of working on the tasks were video-recorded, and we also took notes during the observation and asked the participants to provide more details after each task. We employed a bottom-up approach of thematic coding (Braun & Clarke, 2006) to analyze the qualitative data collected. Two investigators coded the data together to ensure trustworthiness.

5.2. Findings

In general, participants appreciated that their browsing history was organized and visualized for them. Compared to the traditional history function in most browsers, our system helped users develop a better awareness of the time they spent browsing various topics on the web. Participants found the interface design to be aesthetically pleasing and enjoyable. "It looks really nice. I like the combination of colors and layouts," stated by one participant. They enjoyed seeing an overview of their browsing patterns, and valued the multiple interactions that were provided. As one participant said, "It is much easier to tell how much time I am spending on any given topic. I did not realize how long I'm spending on certain things on a traditional history [function]." Compared with the traditional history function, participants especially liked the single-day view. "I like that I can choose a specific day and focus on a specific day," one participant said. The visualizations that participants liked best were the wordCloud, ThemeRiver, and the URL list; they also commented that all the visualizations were useful in their own right. Compared to Chrome's history functions, the participants all considered the visualizations to offer significant advantages, especially for the following:

- identifying different topics and trends
- determining time spent browsing different topics
- finding previously visited web pages
- identifying useful keywords (via the wordCloud)

- narrowing down potential candidates in URL list using the coordinated views
- recalling forgotten topics by viewing keywords
- identifying most popular topics and browsing days/times

As mentioned previously, for Tasks 1 to 5, we asked participants to compare our system to Chrome's history function, and rank each system on a Likert scale from 1 to 7 (from very hard to very easy). Table 2 shows the average Likert scale scores. Apparently, participants prefer our design much more than the standard history function in Chrome for the assigned tasks. Table 3 shows results recorded from the study. We can see that participants can glean considerable information about their browsing history as a result of using our system.

Table 2. Average Scores on the Five Tasks.

Task	Personal Web Library	History function in Chrome
1	6.5	1.83
2	6.83	2.83
3	6.67	3.00
4	6.83	3.33
5	6.33	1.83

Table 3. Results on the Five Tasks.

Task	Recall without PWL	Recognition with PWL
1	Participants identified 5, 3, 5, 2, 4, 3 topics respectively. 4 participants listed specific websites	Participants recognized most topics—especially those with more frequent keywords
2	Participants had a rough idea of topics at large time granularities	Participants easily identified temporal patterns in the ThemeRiver
3	Participants recalled a topic with little detail. 4 participants could recall the website	Participants discussed the topic and context in much more detail
4	Participants could remember multiple visits within 3 days, but not many beyond that	Participants identified most topics and temporal patterns, even from distant days
5	4 participants found the page, 2 did not. Participants chose a starting time, then scrolled through the history log to examine each page they visited	Participants were able to find the page; they all used both the wordCloud and ThemeRiver, which greatly reduced the search space

Aside from the positive comments, participants pointed out some features which needed to be improved. The main problem that was identified relates to the ThemeRiver chart. Participants generally understood what a ThemeRiver communicates, such as the temporal patterns, topics, and rough number of visits. However, some participants were not sure how to read specific values from the ThemeRiver, and were unsure whether specific values were encoded using height or area. The text on the ThemeRiver was difficult to read, especially for inexperienced users. The filled-area chart under the ThemeRiver also required some time to be understood when people saw it for the first time. Another problem reported by many participants was that bars on the single-day density chart were too small to hover over or click. A possible future improvement is to use a magic lens such as a fish-eye view.

Based on the post questionnaire, the overall satisfaction measured by the participants was between "almost satisfied" and "very satisfied." Participants noted that the clearest parts of the design were the interface layout, structure, and linked interactions. Participants found it very easy to explore browsing patterns. However, some participants also said that they did not like the keyword group containing miscellaneous topics. Two participants thought there was a small learning curve when they first saw the visualization, but most participants thought that—as one participant put it—"it is not hard to learn at all." In terms of flexibility, participants asked for a button to go back to the main

page. The small bars on the density chart and the thin areas on the ThemeRiver chart were the main features that participants thought could still be improved.

5.3. Discussion

Our system proposes a new way of organizing and presenting a user's web browsing history. We perceive there to be three broad contributions to our work. First, we devised an effective method for organizing browsing history based on two dimensions—topics and time. By clustering keywords, webpages can be grouped by topics, which carry significant temporal patterns. Second, we proposed a usable and pleasing visualization system that helps users view topics and temporal patterns, allowing them to become aware of their personal web browsing behavior. We described the design process that we followed, including major decisions with supporting rationale. Third, we identified a number of tasks that users reported to be better facilitated with our interactive visualization interface. Researchers and designers can build on these findings to conduct further studies or to implement new visualization tools.

One specific contribution of this research is the new method of supporting revisitation behavior. This new method, unlike the searching method provided in common web browsers, reconstructs the visiting context to help users recall or locate items of interest. Won et al. (Won et al., 2009a) found that many users can "use the memory of activities performed before or after the visit to remember the visit itself". The evaluation of our system suggests that visualizing visit counts and grouping distribution along a timeline can help users recall or relocate visits in the past with context. When users have multiple memory hints about a target topic or website in different time slots, the provided context can help locate relevant items in memory. The cognitive design principle that "recognition is better than recall" served as one core principle of our system design. It is easy for users to recognize related words, highlight the group in the time-oriented visualization, and locate the exact webpages visited. Our "Personal Web Library" integrates different visualizations together to reconstruct the browsing context, which turns out to be supportive for understanding and making use of web browsing history.

From a user experience design perspective, the system has achieved its design goals. The coding scheme and visualization design proved to be intuitive and self-explanatory, as participants were able to understand and describe the meaning of all sections and use them fluently without guidance. Most participants agreed that the consistency of the system interface enhanced its ease of use and learnability. According to participants answers, the interface layout and overall system structure is clear and easy to understand. Participants were always aware of the current system status by seeing the highlighted day in the density chart and the title of the view. Users could also interact with the system with great flexibility due to the coordinated views and brushing and linking interactions. According to the post-experiment questionnaire answers, although much information was provided, participants did not find the visualizations to be overwhelming. Participants enjoyed the different levels of information provided in the simple two-interface structure.

5.4 Key Takeaways

Here we list what we believe are the key takeaways from this work. Although these have been discussed elsewhere in the paper, this section can serve as a summary and quick reference guide.

- Users appreciate being able to see their browsing history, and suggest they are interested in improving their time management skills based on an enhanced awareness of their own browsing habits. Our work can contribute useful ideas to researchers and designers interested in personal informatics, quantified self, and other similar areas.
- We have identified a set of tasks that can be used to study tools that visualize browsing history. These can be used to develop new interactive visualization techniques, to evaluate existing tools, and to set up experiments and usability studies.
- Through both our formative feedback during development and our usability testing, we have identified heuristics for appropriate time granularities when visualizing browsing

history. We found that 6 hours seems to be the best interval for the 2-week period. The 6-hour granularity naturally divides a day into morning, afternoon, evening, and night. These chunks are easily mapped to typical daily activities such as work, leisure, and sleep, which seems to be in line with the way users mentally categorize parts of the day that cannot be immediately recalled from memory. We also found that two main interface levels worked well for users—namely, a two-week view and a single-day view. These findings can be used as general heuristics when designing new tools and conducting studies.

5.5 Future Work

Although PWL provides an intuitive way for users to understand, navigate, and make use of their web browsing history, there is still further development to be done. First, we may incorporate better text analysis algorithms to extract keywords from the page content instead of from the title and metatags. Such extracted keywords may better describe webpages and result in better grouping results. Second, finding or creating a more accurate clustering algorithm to organize keywords into more suitable groups may be beneficial. From the visualization and evaluation, we can see that our keyword extracting method and the k-means algorithm can successfully group keywords for participants to recognize their topics of interest. However, after examining each user's particular topics, it is clear that our data processing and clustering approach can be improved. Related keywords are sometimes separated into two different clusters, and unrelated keywords are sometimes clustered together. Although the main focus of this work is not on clustering and topic extraction per se, future work can implement better algorithms to improve the overall functionality of our system. Third, adding more interactive features where appropriate, such as magic lenses to improve the single-day view, will likely improve the value of our system.

6. Conclusion

In this paper, we proposed a new approach to organizing and visualizing users' web browsing history, with the aim of helping users gain insights into and improve their web browsing experience. The design of our "Personal Web Library" shows a user's topics in groups, with time distributions, at different levels of time granularity. We described our design process, from identifying requirements through design ideation, iterating on early prototypes, and settling on a final prototype. Based on an analysis of the benefits of different visualizations, we chose clustered wordClouds, ThemeRiver charts, filled-area plots, stacked bar charts, and URL lists. The visualizations in our system are coordinated in multiple views that are linked via interaction techniques. Users can easily see patterns and trends at different time granularities, recall pages from the past, and understand the local context of a browsing session. User testing showed that our system is usable and attractive. Its flexibility provides users with much more information than the traditional history function in modern web browsers. Participants in our study gained an improved awareness of their web browsing patterns. Participants mentioned that they were willing to improve their time management after viewing their browsing patterns. As more and more daily activities rely on the internet and web browsers, browsing data captures a large part of users' lives. Providing users with interactive visualizations of their browsing history can facilitate personal information management, time management, and other meta-level activities. We hope the work we conducted can spur more research contributions in this underdeveloped yet important area.

References

Abrams, D., Baecker, R., & Chignell, M. (1998). Information archiving with bookmarks: personal

Web space construction and organization. In *Proceedings of the SIGCHI Conference on*

Human Factors in Computing Systems (pp. 41–48). ACM Press/Addison-Wesley Publishing Co. Retrieved from http://dl.acm.org/citation.cfm?id=274651

Adar, E., Teevan, J., & Dumais, S. T. (2008). Large scale analysis of web revisitation patterns. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 1197–1206). ACM. Retrieved from http://dl.acm.org/citation.cfm?id=1357241

Ayers, E. Z., & Stasko, J. T. (1995). *Using graphic history in browsing the World Wide Web*. Georgia Institute of Technology. Retrieved from https://smartech.gatech.edu/handle/1853/3557

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of Machine Learning Research*, *3*(Jan), 993–1022.

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, *3*(2), 77–101.

Cernea, D., Truderung, I., Kerren, A., & Ebert, A. (2013a). WebComets: A Tab-Oriented Approach for Browser History Visualization. In *GRAPP/IVAPP* (pp. 439–450). Citeseer. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.302.7649&rep=rep1&type=pdf

Cernea, D., Truderung, I., Kerren, A., & Ebert, A. (2013b). WebComets: A Tab-Oriented Approach for Browser History Visualization. In *GRAPP/IVAPP* (pp. 439–450).

Chtivelband, I. (2012). *HistoryLane: Web Browser History Visualization Method*. Retrieved from http://www.diva-portal.org/smash/record.jsf?pid=diva2:833178

Collins, C., Viegas, F. B., & Wattenberg, M. (2009). Parallel tag clouds to explore and analyze faceted text corpora. In *Visual Analytics Science and Technology, 2009. VAST 2009. IEEE Symposium on* (pp. 91–98). IEEE. Retrieved from http://ieeexplore.ieee.org/abstract/document/5333443/

Cugini, J., & Scholtz, J. (1999). VISVIP: 3D visualization of paths through web sites. In *Database and Expert Systems Applications, 1999. Proceedings. Tenth International Workshop on* (pp. 259–263). IEEE. Retrieved from http://ieeexplore.ieee.org/abstract/document/795175/

Cui, W., Liu, S., Tan, L., Shi, C., Song, Y., Gao, Z., … Tong, X. (2011). Textflow: Towards better understanding of evolving topics in text. *IEEE Transactions on Visualization and Computer Graphics*, *17*(12), 2412–2421.

Dennis, S., Landauer, T., Kintsch, W., & Quesada, J. (2003). Introduction to latent semantic analysis. In *Slides from the tutorial given at the 25th Annual Meeting of the Cognitive Science Society, Boston*. Retrieved from https://pdfs.semanticscholar.org/beee/452ebf1d92cf13879cc8422737f12b65cade.pdf

Dou, W., Wang, X., Skau, D., Ribarsky, W., & Zhou, M. X. (2012). Leadline: Interactive visual analysis of text data through event identification and exploration. In *Visual Analytics Science and Technology (VAST), 2012 IEEE Conference on* (pp. 93–102). IEEE. Retrieved from http://ieeexplore.ieee.org/abstract/document/6400485/

Endert, A., Fiaux, P., & North, C. (2012). Semantic interaction for visual text analytics. In *Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 473–482). ACM. Retrieved from http://dl.acm.org/citation.cfm?id=2207741

Ericsson, K. A., & Simon, H. A. (1998). How to study thinking in everyday life: Contrasting think-aloud protocols with descriptions and explanations of thinking. *Mind, Culture, and Activity*, *5*(3), 178–186.

Fisher, B., Agelidis, M., Dill, J., Tan, P., Collaud, G., & Jones, C. (1997). CZWeb: Fish-eye views for visualizing the world-wide web. *Advances in Human Factors Ergonomics*, *21*, 719–722.

Fulda, J., Brehmer, M., & Munzner, T. (2016). TimeLineCurator: Interactive authoring of visual timelines from unstructured text. *IEEE Transactions on Visualization and Computer Graphics*, *22*(1), 300–309.

Gandhi, R., Kumar, G., Bederson, B., & Shneiderman, B. (2000). Domain name based visualization of Web histories in a zoomable user interface. In *11th International Workshop on Database and Expert Systems Applications, 2000. Proceedings* (pp. 591–598). https://doi.org/10.1109/DEXA.2000.875085

Haggstrom, D. A., Saleem, J. J., Russ, A. L., Jones, J., Russell, S. A., & Chumbler, N. R. (2011). Lessons learned from usability testing of the VA's personal health record. *Journal of the American Medical Informatics Association*, *18*(Supplement 1), i13–i17.

Havre, S., Hetzler, B., & Nowell, L. (2000). ThemeRiver: Visualizing theme changes over time. In *Information Visualization, 2000. InfoVis 2000. IEEE Symposium on* (pp. 115–123). IEEE. Retrieved from http://ieeexplore.ieee.org/abstract/document/885098/

Keim, D. A. (2002). Information visualization and visual data mining. *IEEE Transactions on Visualization and Computer Graphics*, *8*(1), 1–8.

Kodinariya, T. M., & Makwana, P. R. (2013). Review on determining number of Cluster in K-Means Clustering. *International Journal*, *1*(6), 90–95.

Kumar, R., & Sharan, A. (2014). Personalized web search using browsing history and domain knowledge. In *Issues and Challenges in Intelligent Computing Techniques (ICICT), 2014 International Conference on* (pp. 493–497). IEEE. Retrieved from http://ieeexplore.ieee.org/abstract/document/6781332/

Lam, H., Bertini, E., Isenberg, P., Plaisant, C., & Carpendale, S. (2012). Empirical Studies in Information Visualization: Seven Scenarios. *IEEE Transactions on Visualization and Computer Graphics*, *18*(9), 1520–1536. https://doi.org/10.1109/TVCG.2011.279

Larkin, J. H., & Simon, H. A. (1987). Why a Diagram is (Sometimes) Worth 10,000 Words. *Cognitive Science*, *11*, 65–99.

Li, H., & Yamanishi, K. (2003). Topic analysis using a finite mixture model. *Information Processing & Management*, *39*(4), 521–541.

Matsuo, Y., & Ishizuka, M. (2004). Keyword extraction from a single document using word co-occurrence statistical information. *International Journal on Artificial Intelligence Tools*, *13*(01), 157–169.

Matthijs, N., & Radlinski, F. (2011). Personalizing web search using long term browsing history. In *Proceedings of the Fourth ACM International Conference on Web Search and Data Mining* (pp. 25–34). ACM. Retrieved from http://dl.acm.org/citation.cfm?id=1935840

Mayer, M. (2009). Web History Tools and Revisitation Support: A Survey of Existing Approaches and Directions. *Found. Trends Hum.-Comput. Interact.*, *2*(3), 173–278. https://doi.org/10.1561/1100000011

Milic-Frayling, N., Sommerer, R., & Rodden, K. (2003). WebScout: Support for revisitation of web pages within a navigation session. In *Web Intelligence, 2003. WI 2003. Proceedings. IEEE/WIC International Conference on* (pp. 689–693). IEEE. Retrieved from http://ieeexplore.ieee.org/abstract/document/1241297/

Munzner, T. (2009). A Nested Model for Visualization Design and Validation. *IEEE Transactions on Visualization and Computer Graphics*, *15*(6), 921–928. https://doi.org/10.1109/TVCG.2009.111

Purcell, A. T., & Gero, J. S. (1998). Drawings and the design process: A review of protocol studies in design and other disciplines and related research in cognitive psychology. *Design Studies*, *19*(4), 389–430. https://doi.org/10.1016/S0142-694X(98)00015-5

Rind, A., Aigner, W., Wagner, M., Miksch, S., & Lammarsch, T. (2016). Task Cube: A three-dimensional conceptual space of user tasks in visualization design and evaluation. *Information Visualization*, *15*(4), 288–300.

Rivadeneira, A. W., Gruen, D. M., Muller, M. J., & Millen, D. R. (2007). Getting our head in the clouds: toward evaluation studies of tagclouds (p. 995). ACM Press. https://doi.org/10.1145/1240624.1240775

Roberts, J. C. (2007). State of the art: Coordinated & multiple views in exploratory visualization. In *Coordinated and Multiple Views in Exploratory Visualization, 2007. CMV'07. Fifth International Conference on* (pp. 61–71). IEEE. Retrieved from http://ieeexplore.ieee.org/abstract/document/4269947/

Scholtz, J., Plaisant, C., Whiting, M., & Grinstein, G. (2014). Evaluation of visual analytics environments: The road to the Visual Analytics Science and Technology challenge evaluation methodology. *Information Visualization*, *13*(4), 326–335.

Sedig, K., & Parsons, P. (2016). Design of Visualizations for Human-Information Interaction: A Pattern-Based Framework. *Synthesis Lectures on Visualization*, *4*(1), 1–185.

Shneiderman, B. (1996). The eyes have it: A task by data type taxonomy for information visualizations. In *Visual Languages, 1996. Proceedings., IEEE Symposium on* (pp. 336–343). Boulder, CO: IEEE. Retrieved from http://ieeexplore.ieee.org/xpls/abs_all.jsp?arnumber=545307

Steinbach, M., Karypis, G., Kumar, V., & others. (2000). A comparison of document clustering techniques. In *KDD workshop on text mining* (Vol. 400, pp. 525–526). Boston. Retrieved from https://pdfs.semanticscholar.org/c110/0f525044b2b926f7bd7f407ce7b0157bcfd8.pdf

Tauscher, L. M. (1996). Evaluating history mechanisms: An empirical study of reuse patterns in World Wide Web navigation. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.43.5856

Van Kleek, M., Moore, B., Xu, C., & Karger, D. R. (2010a). Eyebrowse: Real-time Web Activity Sharing and Visualization. In *CHI '10 Extended Abstracts on Human Factors in Computing Systems* (pp. 3643–3648). New York, NY, USA: ACM. https://doi.org/10.1145/1753846.1754032

Van Kleek, M., Moore, B., Xu, C., & Karger, D. R. (2010b). Eyebrowse: Real-time web activity sharing and visualization. In *CHI'10 Extended Abstracts on Human Factors in Computing Systems* (pp. 3643–3648). ACM. Retrieved from http://dl.acm.org/citation.cfm?id=1754032

Wartena, C., & Brussee, R. (2008). Topic detection by clustering keywords. In *Database and Expert Systems Application, 2008. DEXA'08. 19th International Workshop on* (pp. 54–58). IEEE. Retrieved from http://ieeexplore.ieee.org/abstract/document/4624691/

Wei, F., Liu, S., Song, Y., Pan, S., Zhou, M. X., Qian, W., … Zhang, Q. (2010). Tiara: a visual exploratory text analytic system. In *Proceedings of the 16th ACM SIGKDD international conference on Knowledge discovery and data mining* (pp. 153–162). ACM. Retrieved from http://dl.acm.org/citation.cfm?id=1835827

Wexelblat, A., & Maes, P. (1999a). Footprints: history-rich tools for information foraging. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems* (pp. 270–277). ACM. Retrieved from http://dl.acm.org/citation.cfm?id=303060

Wexelblat, A., & Maes, P. (1999b). Footprints: History-rich Tools for Information Foraging. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 270–277). New York, NY, USA: ACM. https://doi.org/10.1145/302979.303060

Won, S. S., Jin, J., & Hong, J. I. (2009a). Contextual web history: using visual and contextual cues to improve web browser history. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 1457–1466). ACM. Retrieved from http://dl.acm.org/citation.cfm?id=1518922

Won, S. S., Jin, J., & Hong, J. I. (2009b). Contextual web history: using visual and contextual cues to improve web browser history. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (pp. 1457–1466). ACM. Retrieved from http://dl.acm.org/citation.cfm?id=1518922

Yu, W., & Ingalls, T. (2011). Trails–An Interactive Web History Visualization and Tagging Tool. In A. Marcus (Ed.), *Design, User Experience, and Usability. Theory, Methods, Tools and Practice* (pp. 77–86). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-21708-1_10

Zu Eissen, S. M., & Stein, B. (2002). Analysis of clustering algorithms for web-based search. In *International Conference on Practical Aspects of Knowledge Management* (pp. 168–178). Springer. Retrieved from http://link.springer.com/chapter/10.1007/3-540-36277-0_16