Master Thesis
Software Engineering
Thesis no: MSE-2012-106
09 2012

HistoryLane : Web Browser History Visualization Method

Igor Chtivelband

School of Computing
Blekinge Institute of Technology
SE-371 79 Karlskrona
Sweden

This thesis is submitted to the School of Computing at Blekinge Institute of Technology in partial fulfillment of the requirements for the degree of Master of Science in Software Engineering. The thesis is equivalent to 20 weeks of full time studies.

Contact Information:
Author(s):
Igor Chtivelband 820109P892
E-mail: igor.chtivelband@gmail.com

BTH University advisor(s):
Dr. Mikael Svahnberg
School of Computing

TUK University advisor(s):
Prof. Achim Ebert
Dipl.-Ing. Daniel Cernea

School of Computing
Blekinge Institute of Technology
SE-371 79 Karlskrona
Sweden

Internet : www.bth.se/com
Phone : +46 455 38 50 00
Fax : +46 455 38 50 57

Abstract

With the evolution of Internet, users exploit it increasingly intensively for achieving their goals: sending emails, playing games, watching videos, reading news. They do it through Web browsers, that may vary in exact implementation, but have similar core functionalities. One of these core functionalities is the access to a browsing history. However, as browsing patterns are getting more complex, the traditional history tools become insufficient. Visualization of browsing history might be helpful in that case.

In this study we propose a novel approach for browsing history visualization, named HistoryLane, which fits the parallel browsing paradigm, common for modern browsers. The main goal of HistoryLane is enabling the user to gain insight into his own or into other users' parallel browsing patterns over time.

Principles of HistoryLane visualization approach are formulated based on recommendations, found during structured literature review. These principles constitute the base for a prototype, which was implemented as a Firefox extension. To evaluate the effectiveness of HistoryLane we conducted a survey and a quantitative experiment.

The results of the evaluation show that HistoryLane is perceived by users as effective and intuitive method for browsing history visualization.

Keywords: Parallel browsing behavior, tab-based visualization

Acknowledgments

First of all I want to thank my direct supervisor Mr. Daniel Cernea for his knowledge and experience that he generously shared with me from the beginning until the end of this master thesis. He always found time and patient to guide me when I needed that. Also I am very thankful to Dr. Mikael Svahnberg (BTH) for his valuable feedback and to Prof. Achim Ebert (TU Kaiserslautern) for his readiness to evaluate this work.

The experimental part of this thesis would not be completed without help of numerous volunteers who contributed their time to survey and experiment participation, so I would like to express my deep gratefulness to those people.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Problem Statement

About 2.1 billion of Internet users were recorded worldwide in 2011 [24]. It means that a third of Earth population uses Web browsers, which makes them one of the most popular categories of software products. Not only the number of users is growing, but also the variety of tasks that they want to complete using Web browsers. People send emails, play on-line games, stream videos, read news, etc.

Until recently the main browsing paradigm involved visiting the sequence of Web pages in the same browser window. However since introduction of browser tabs, browsers started to support the parallel browsing paradigm, allowing users to switch between tabs in the same window [10]. As a result users visit more Web sites and perform more actions, making the analysis of those actions a truly difficult task.

There are multiple reasons for such analysis: users themselves do not remember where they have read an interesting article that they want to post in a social network (80% of Web sites are revisited [2], but it is not always trivial to revisit, improved revisitation support could make the daily work experience of billions of users easier [20]). Managers may be interested in checking what Web sites are popular among company employees, anthropologists may use such data for investigating the behavior patterns of Web surfers.

Most of modern browsers store data about visited Web pages and present that data to users on request. List of collected parameters usually includes title, url, timestamp, and sometimes thumbnail images [32]. These history entries are shown in a chronological order or based on the frequency of the visits. All modern browsers in addition index textual context of the visited Web pages for provision of textual search later. However there are multiple problems with existing representations of browsing history.

1. They do not provide information about how the Web site was reached (e.g. through search engine, hyperlink or typing URL) [18].

2. They do not record what was the amount of user's activity or idleness at specific Web page, despite that this data may be very useful for building

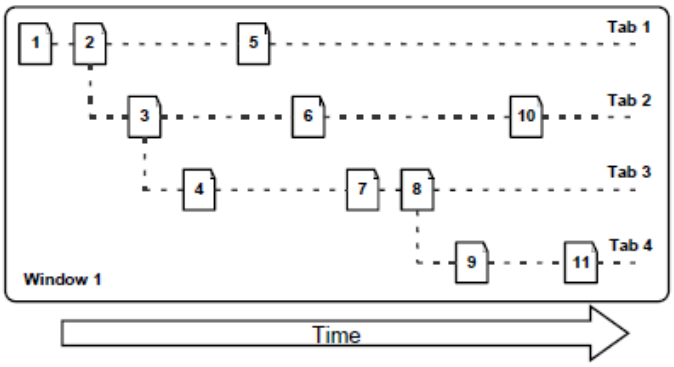


Figure 1.1: An example of parallel browsing session visualization [10]

user profile later [15].

3. Behavioral patterns are hard to extract [18].

4. Most of modern browsers allow parallel browsing using multiple open tabs, however existing solutions for browsing history visualization do not provide ability to map accessed Web sites to relevant open tabs. Such visualization may be very intuitive and simple, as well as useful for discerning various browsing habits of users, see Figure 1.1 [12].

5. Accessed Web sites are not summarized/categorized. Aggregation of data is important, because displaying raw data about large numbers of URLs is not efficient [25].

1.2 Scope of Thesis Work

The aim of this master thesis is to formulate and evaluate a new approach for intuitive and effective visualization of Web browsing history that matches the modern parallel browsing paradigm. This aim is achieved by taking the following steps:

- Conduction of structured literature review, in order to investigate number of aspects:
  - Gap between existing browsing history tools and requirements formulated by other researches
  - Previous attempts to visualize browsing history using graphical entities
  - Cognitive perception of Web browsing process

  - Parallel browsing behavior patterns

- Formulation of innovative browsing history visualization principles and guidelines
- Creation of a working prototype that demonstrates formulated ideas
- Evaluation of the effectiveness and intuitiveness of new visualization approach using the prototype

The working prototype that is mentioned above is created only for demonstration needs and may lack usability and reliability that are typical for end-user aimed solutions.

1.3 Research Questions

In order to achieve the goals that were defined in the previous section, the following research questions should be addressed:

1. RQ1: Does the aggregation of browsing history into tab sections makes the perception of this history more convenient for users?
2. RQ2: How to visualize collected history data?
3. RQ3: What are the benefits of a browsing history visualization?

1.4 Research Methodologies

The first step in answering the research questions consists of conducting a detailed and comprehensive literature review. Based on findings from the literature review we formulate the main guidelines for new browsing history visualization approach. In order to demonstrate those ideas a working prototype is created. Later this prototype is installed on volunteers' computers to make them familiar with a new tool. Then both qualitative (survey) and quantitative (tasks) experiments are conducted using help of volunteers. As a last step experiments' data is analyzed using statistical tools and answers for research questions are presented (see Figure 1.2).

1.5 Outline of the Thesis

The rest of this thesis report is structured in following way: Chapter 2 presents the background about World Wide Web, browsers and previous attempts to visualize browsing history. Chapter 3 contains principles of history visualization that

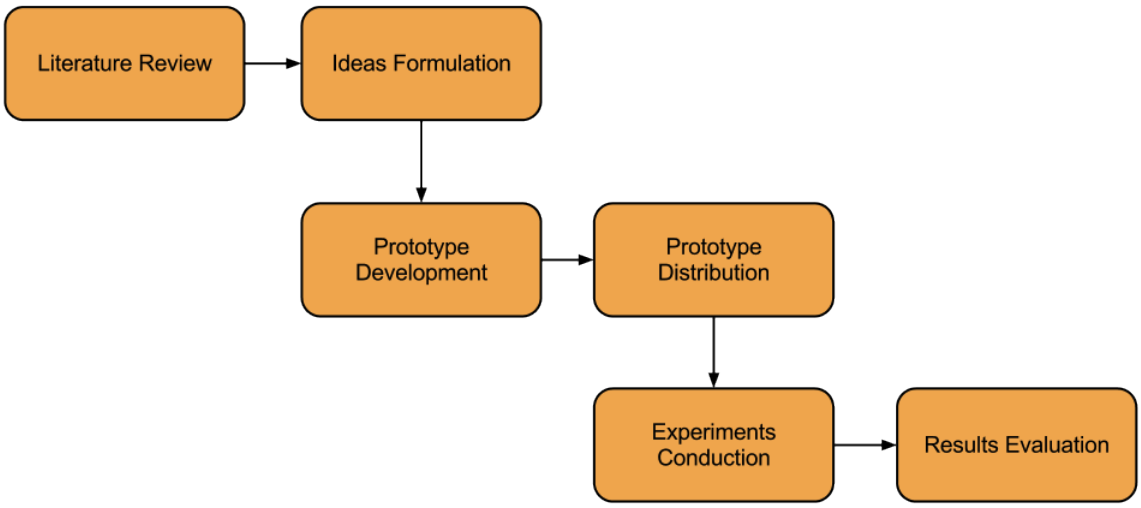


Figure 1.2: Schematic diagram of the research methodology

we formulated based on literature review and our own ideas. In Chapter 4 we describe the prototype that was created applying formulated visualization principles. Chapter 5 presents the prototype evaluation process. The results of this evaluation are discussed in Chapter 6. In Chapter 7 we present the conclusions of the entire thesis work and propose in what directions research may be continued. Finally in Appendix section we provide additional materials of conducted experiments.

Chapter 2

Background and related work

This chapter provides general background about World Wide Web, browsers as technical instruments for accessing Web pages and what history viewing tools these browsers support. In addition previous attempts to visualize browsing history are presented. Their description includes strengths, weakness and comparison of their approaches to the one that is developed within the scope of this master thesis.

2.1 Background

2.1.1 Evolution of World Wide Web

Let us start with defining what Internet and World Wide Web are. Some people may say that they are synonymous, while others will argue, that these are completely different things, because while Internet is "A collection of computer networks based of specific set of network standards, namely, TCP/IP" [6], the World Wide Web is "The universe of of network-accessible information, an embodiment of human knowledge" [4]. Since for majority of users the only experience with the Internet is using World Wide Web, let us assume in scope of this master thesis that WWW and Internet are synonymous.

The roots of WWW go back to 1969 when Defense Advanced Research Projects Agency (DARPA) established an early internetwork called ARPANET, the Advanced Research Projects Agency Network, that connected all research centers to facilitate data exchange. In 1979 Standard Generalized Markup Language (SGML) was invented to enable sharing of documents for large projects by separating contact from the presentation layout, making possible to parse same document in different manners. In 1989 Tim Berner-Lee, trying to improve the documentation handling and sharing in CERN, developed Networked Hypertext protocol. At that time CERN was connected to Internet for over 2 years, but scientists in CERN were looking for better approach for circulating of their publications and information among the research world [7].

In a couple of years Tim Berner-Lee developed the initial software for hypertext server programming and made it available for free download. This paved

the way for the popularity of the Web. Berners-Lee called his hypertext based system World Wide Web and very fast it became extremely popular among the scientific community. However the main problem of Web was the lack of software to read the hypertext documents, Web browsers were required to continue Web evolution and soon they appeared [7].

2.1.2 Web Browsers: Then and Now

In 1993 Marc Andressen and his team from University of Illinois created a program called Mosaic that could render a hypertext document and interpret its contents, so they could be displayed on the user's screen in a graphical format. This program, later declared as a first Web browser, opened the gates of Web for the general public. Mosaic was also distributed for free, fact which definitely contributed to its popularity [7].

Soon Marc Andressen started his own company named Netscape. This new company in 1994 released Netscape Navigator which became the most used Web browser of that time by reaching 90% market share at the peak of its popularity [30].

A Software giant like Microsoft could not stay aside and not participate in the development of Web browsers. In 1995 Microsoft released the first version of Internet Explorer (IE) which was included in Windows 95 operating systems. Due to outstanding success of Windows 95, IE browser became an extremely popular browser with 85% market share at 2002 [30].

In 1996, after working with Telenor, Opera releases its own Web browser named Opera 2.0. Market share of Opera browsers never exceeded 3% barrier in last 16 years [30].

Another software giant, Apple, launched a Web browser for Mac Os X operating system on June 2003. It was named Safari and was included as default browser in Mac OS X 10.3. Up to nowadays amount of Safari users stays under 5% of total amount of Web users [30].

Blake Ross released Firefox 1.0 at 2004. The origins of Firefox project are in Mozilla open source project, which was started by Netscape at 1998. Firefox continues to be one of the most popular modern browsers, however the peak of its popularity was in 2009 with 48% market share [30].

After Microsoft and Apple, Google also released its own browser named Chrome in 2008. Since then there is a continuous growth in number of users, which currently stands on 43% [30].

According to updated statical data (July 2012), most popular Web browsers are: Chrome (42.9%), Firefox (33.7%), IE (16.3%), Safari (3.9 %) and Opera (2.1%). See Figure 2.1

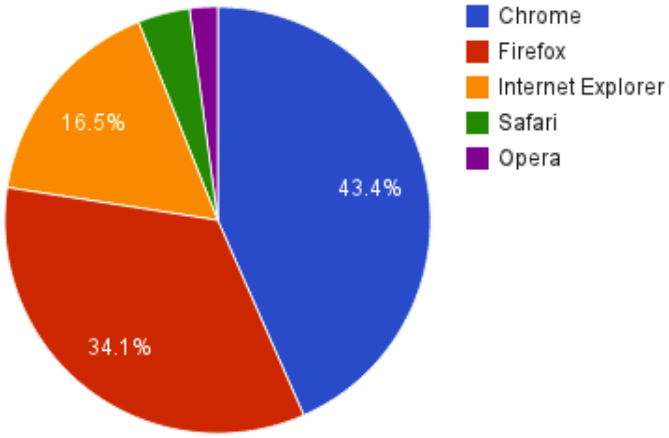


Figure 2.1: Market shares of modern browsers in July 2012

2.1.3 History Tools in Modern Web Browsers

In Internet Explorer users can see the list the visited Web pages sorted by date, popularity or aggregated by some time period. In addition there is an option to perform a search by keywords. The results are always presented as list of Web page titles.

In Google Chrome, history is represented as a list of page titles accompanied by the exact time stamp of last visit and favicon (Favorite Icon). The history menu opens in a new tab, similar to a regular Web page. Search, based on key words, is available as well. In addition Chrome provides a visualization of most visited Web sites, when user opens a new tab, where Web sites are exhibited as thumbnail images.

In Firefox, history is a list of entries aggregated by time interval, where each entry includes page title, favicon and URL. Search capability is limited to URL of Web pages, while in Chrome it indexes the context of pages as well.

Apple Safari 5 browsing history includes both textual links and a slick Top Sites page [12]. Top Sites page includes 3D style mart with thumbnails of most visited Web pages. Those thumbnails may convey additional information, like indication if content of specific Web page was changed since the last visit.

History tools in modern browsers have few common problems:

- All list entries have similar appearance, which doesn't depend on category of Web site.
- There is no connection between related page visits. User can not reconstruct pattern of his browsing.
- History entries lists are hidden in browser menu, thus they are used very rarely by users. Some users even don't know about their existence [27].

Visualization approach, that we developed as part of this master thesis, uses another strategy to present users' history. It is based on assumption that users remember their browsing paths as they remember geographical paths, based on visual anchors, and on visual associations between Web site category and specific color.

2.2 Related Work

Results of experiments show, that use of history mechanism may have significant effect on user satisfaction and performance when revisiting Web pages [22]. In addition they show that use of visual aids in history mechanisms is more effective than the use of textual only data [22]. In another set of experiments Mascoet [18] found that users, trying to revisit specific Web page make 50% less "mistakes" when they use the graphical visualization, comparing to case when they use textual representation only. These findings demonstrate that it is important to develop and/or enhance history visualization mechanisms. Not surprisingly many researches tried to do it before us. Here is a short overview of their attempts.

2.2.1 Time-line Based Visualizations

isoBrowser (Figure 2.2) was developed by Hodgkinson [9] as part of his master thesis. It presents the browsing history as a scrollable timeline, which is simple, but effective visualization solution. Timeline consists of thumbnails, which are put in a chronological order. After clicking on thumbnail opens the pop-up window with the detailed information regarding this page (time of first visit, URL, time of last visit). User can drag the thumbnails from timeline, put them "aside" and organize them in stacks. In the latter case, the shadow images of thumbnail remains in timeline scroll. Opposite to folders in operating systems, there is no hierarchy of stacks (you can not create stack inside of stack). The intention of stack use is to remain at one interface level.

Browsing history visualization problem is relevant not only for PCs, but for mobile devices as well. Vartiainen et al. [29] developed a solution called Rolling History for mobile devices that have 4 directions of navigation control and graphics acceleration hardware. The graphical representation of browsing history consists of thumbnails, where each thumbnail presents a Web page, that user has visited. These thumbnails are aligned along a horizontal axis, where currently active Web page is shown as a largest one, making the orientation easier. This solution supports concurrent browsing model, where multiple browser windows can be open simultaneously. Other open browsers are shown as vertical list of thumbnails, where the currently focused one is located in the middle and its history list is shown in horizontal line, see Figure 2.3. If the Web page from the

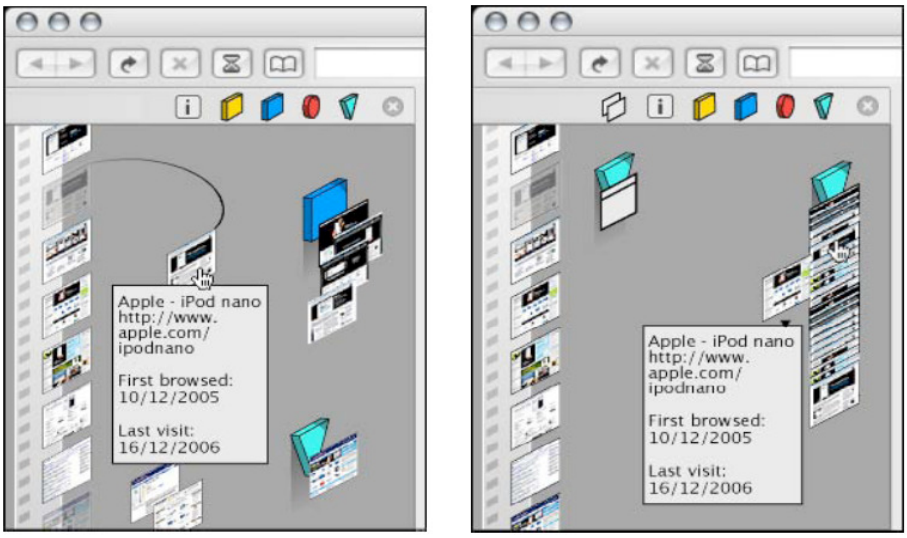


Figure 2.2: isoBrowser. Thumbnails can be organized in stacks for simplicity and order.

history in a horizontal axis is currently open in any window, that window is clearly marked in the window list. The Rolling History is a prominent solution due to effective navigation realization and division of browsing history by windows/tabs.

2.2.2 2D Visualizations

Milic-Frayling et al. [21] deduced that in order to be effective during browsing user has to keep a mental note of both the hierarchical structure and the access sequence of Web pages. To assist users with these tasks they propose to partition the user's navigation into logical sessions, where every session consists of pages, bound by common meaning. New sequence begins with the user's request for specific page by typing the URL or by search action using search engine. These sequences are visualized as a horizontal tree, where new sequence starts a new branch. Branch is shown as a sequence of thumbnails in the order of access as shown in Figure 2.4. As user continues browsing new thumbnails images are appended to current branch. To implement the described visualization Milic-Frayling et al. [21] developed an extension for Internet Explorer named SessionNavigator.

Another session oriented visualization method, named SessionGraphs was proposed by Mayer [20]. SessionGraphs represents visited Web pages as nodes in the directed graph and moves between them as edges. This approach was chosen, because according to Mayer [20] graph's visualization contains more characteristic features than a plain sequential list (changes of direction, loops, etc.). Nodes, that are visited multiple times are visualized only once. A single Web page is

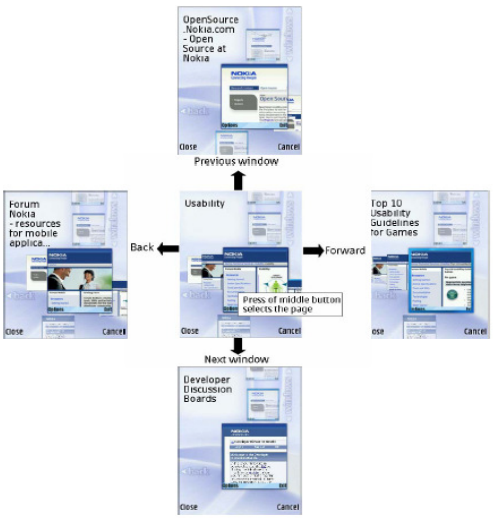


Figure 2.3: Rolling History. Horizontal line contains thumbnails from the history of currently open window, vertical line contains visualization of other open windows.

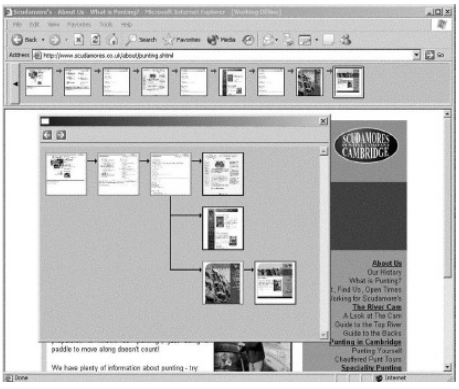


Figure 2.4: SessionNavigator visualized browsing history as horizontal tree [21].

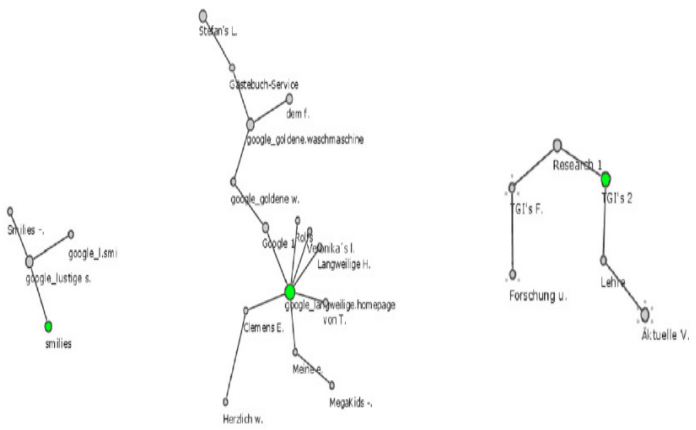


Figure 2.5: SessionGraphs visualization focuses of depicting the shape of entire session [20].

shown as a circle with an attached label as a semantic title. The size of circle depends on the time spent at this Web page. Motion and color are used to highlight the new added and currently visited nodes. The currently visited node is green, while others are gray (see Figure 2.5). User can color nodes manually to create recognizable patterns. Edges can be shown with or without arrows that indicate browsing direction. The main idea of this visualization is to make user perceive the shape of an entire browsing session. To mirror the fluid character of Web activity and to create the playful exploration environment the *fluid surface metaphor* was introduced. The graph that represents the session slowly drift on 2-D surface. Mayer [20] describes it in the following poetical way: *"The behavior should be similar to clusters of sea roses or leaves that drift on a lake's gently moving surface"*. It should attract the user to play with visualization, to manipulate it. Technically this visualization method was implemented as a stand-alone Java application next to Internet Explorer.

Search History Tree (SHT) approach was developed by Simko et al. [26] for providing revisitations support for previously discovered information during search sessions. As the name implies, SHT is a tree-based visualization, where nodes represent user queries. SHT continuously records user activities in browser and constructs a tree-based representation of query modification during sessions. The purpose is to provide an orientation support within history of queries and results. To do so, thumbnail images of visited Web pages are attached to query nodes, see Figure 2.6 .The session is defined based on goals that user wants to achieve rather that instances of Web search application. Thus, the identification of session boundaries is a non-trivial task which is performed using term analysis algorithms.

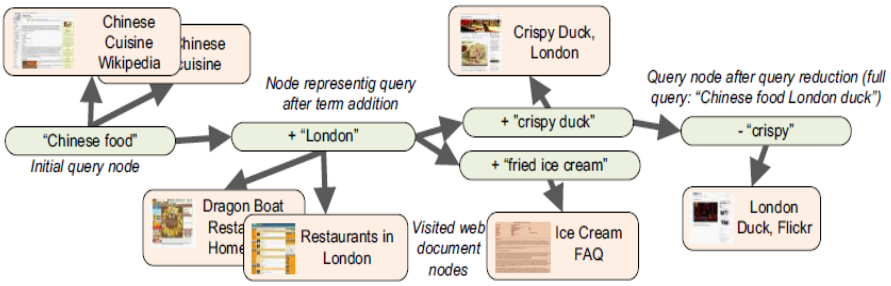


Figure 2.6: An example of search session as it is shown by Search History Tree [26]

2.2.3 3D Visualizations

Frecon and Smith [5] logically concluded that a three-dimensional presentation provides a maximal flexibility in data visualization. WebPath module, that they developed, uses the information contained in the HTML of browsed Web page to produce a representation in 3D space. Each Web page is represented as a cube, labeled by the page's title. Cubes where chosen, because their surfaces can be texture mapped. In addition surface may be used to show images from the Web pages, to increase the recognition. User can manually update the image on the cube with more informative one. To reduce the visual complexity WebPath uses Level of Detail (LOD). The title on the cube is replaced by a simple polygon from a distance. WebPath uses three dimension, thus different metrics may be associated to one of the horizontal axes. The vertical axis is reserved to the time of visit metric. The examples of such metrics for horizontal axes may be the size of the page or the origin server, so that all pages that originate from a single server are aligned. In addition the geographic layout is available, where pages are placed according to the geographical coordinates of their domain registration. It is important to mention, that if Web page is revisited, new cube is generated. Interesting feature of developed solution is a delimiting semi-transparent plane, that can be inserted into the virtual environment and moved along axes, helping to visualize the value of metrics associated with axes.

Another example of system for 3D browsing history visualization was developed by Yamaguchi et al. [34]. They created tool that supports multiple layouts: a book mode, a circle mode and a cube mode (Figure 2.7). Using a book layout users can see each page as if he/she were browsing the book. The book metaphor is widely used in modern system, thus this visualization is intuitive to users. In a circle layout the images of Web pages are placed around the circumference of a circle. In cube layout images of Web pages are put on the surfaces of cube and user can follow the browsing history by rolling over the cube. However, it is not clear, how Yamaguchi et al. [34] deal with big number of images in a cube layout, since the surface area is limited and can not be extended.

Figure 2.7: Different layouts of 3D visualization [34]

2.2.4 Multitab Oriented Visualization

Most of modern popular Web browsers support concurrent browsing using multiple tabs. Multiple tabs are useful for comparing information from multiple on-line sources simultaneously, e.g. for comparing prices. Some Web sites even require the support for multiple tabs by providing essential information in pop-up window [29]. Huang and White [10] found that parallel browsing improves the performance of users by making them work in multitasking mode. This type of browsing has been growing recently and gains more and more popularity. The concurrent browsing behavior requires the appropriate history representation, which would provide a positive and productive user experience [12]. One way of such representation was proposed by Khaksari [12]. He assumes, that an appropriate model for representation of browsing history is a grid, populated with thumbnail images of Web pages. This grid would consist of a number of labeled tabs, where each tab corresponds one-to-one to relevant tab in browser. This history grid resides in the background, a mouse click on history button brings it to the front, by making the browsing display blurred out, see Figure 2.8. A user can easily toggle back and forth between history mode and browsing mode by a History button click. Thumbnail images in the grid can be easily manipulated, zoomed-out, etc. According to Khaksari [12] this method reduces cognitive workload, makes Web browsing experience enjoyable and reduces user's frustration.

2.2.5 Visualizing Categories of Web Pages

Both regular users and behavior analysts are interested to know what types of Web sites are visited by browser users. To do so Web sites should be categorized and single visits should be aggregated according to those categories (news/e-commerce/etc.)

An eminent example of user activities visualization was developed by Reiss and Eddon [25]. They created tool named *webviz* for monitoring the user behavior in real time. *Webviz* gathers data from large number of users, monitoring the URLs that they access, then it summarizes this information by categories of Web sites (provided by OpenDirectory hierarchy) and displays the result. Users can iden-

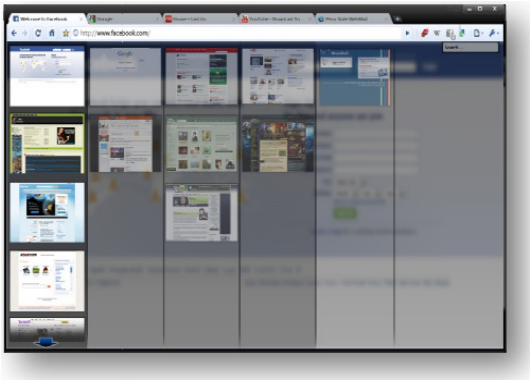


Figure 2.8: Web Browsing History Grid. Vertical columns of thumbnail grid are mapped to corresponding tabs in background [12].

tify browsing patterns, trends, or peaks of unusual activities. Figure 2.9 shows a sample of *webviz* display. Display consists of concentric circles, where each circle represents a different time period. The outermost one is last 5 minutes, the next is previous 5 minutes, next one is last 30 minutes, next the last hour, the prior 2 hours, the last 8 days, etc. The most internal one consists data of 2 last days. These intervals can be configured by user, according to his goals. Each circle is demarcated into multiple categories, that are set in alphabetical order starting 3:00 position. Each category is assigned a color in order to make demarcation more explicit. The examples of categories are "Computers","Sport","Arts". For each category additional information about users, views and number of distinct URLs is available. This information is coded by line, that is visible in the middle of each category. The width of this line indicates the relative number of different users, the frequency of line indicates the relative number of URLs and the amplitude of the line may be used for additional metrics. We find this approach for visualization remarkable, because of non-trivial usage of circles for time periods representation. In addition the usage of colors for categories demarcation is a successful idea also.

Van Kleek et al. [28] assumed that Web browsing trials reflect interests of users and what they do in daily lives. These trails have the potential to help users in various ways, e.g. to keep track of how users spend their time. To exploit this potential Van Kleek et al. [28] developed tool named *Eyebrowse* which provides quick access to the individuals browsing activities and presents trends aggregated by various time intervals. *Eyebrowse* tracks user's activities and generates easy-to-read statistical visualization. For tracking those activities Mozilla Firefox add-on should be installed at user's computer. This add-on collects the data and sends it to remote server. To view his statistical data user browses to *Eyebrowse* Web site. The examples of available statistics are top 25 Web sites by week, frequency and duration of Web browsing activities, daily activity, etc (Figure 2.10).

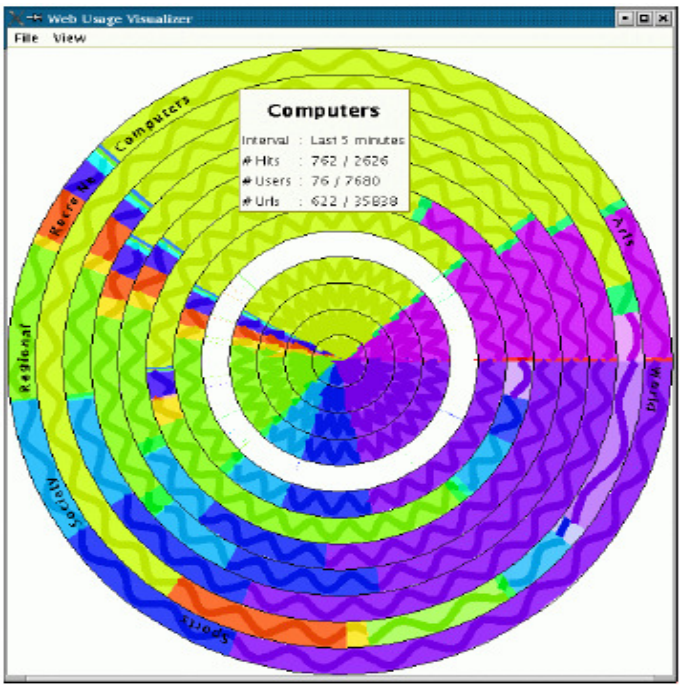


Figure 2.9: The *webviz*. Each circle represents a time interval. Categories within a circle are marked by distinct colors. The brightness of the color and the characteristics of the waving line carry additional information [25].

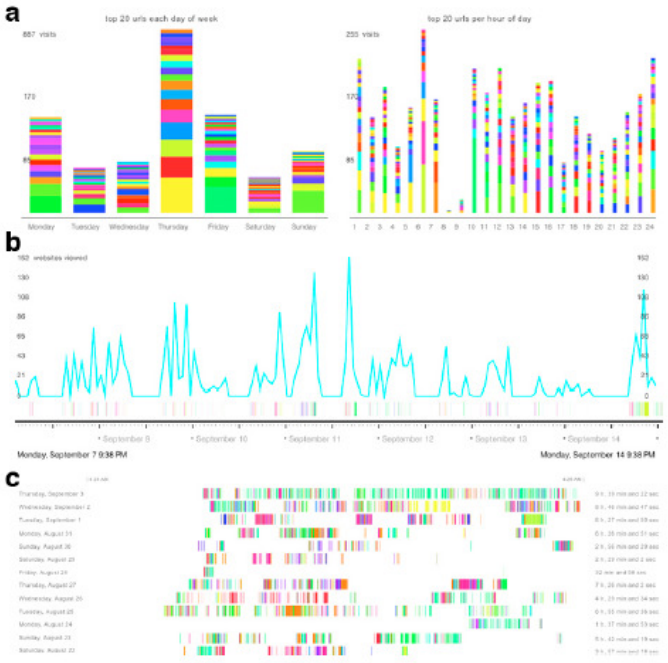


Figure 2.10: *Eyebrowse* shows various data marts (a) Top 20 URLs for day of week and time of day (b) Timeline of pages visited over the course of 1 week (c) Timeline over 20 days [28]

2.3 Comparison of Related Work to History-Lane Approach

Most of visualization methods that were presented above ignore the parallel browsing paradigm that is widely adopted by all modern browsers. Visited Web pages are shown/listed one after another, even if they were accessed simultaneously or almost simultaneously. Another limitation of most of these tools is a lack of ability to provide a general picture from "bird's eye point of view" on last visited Web pages, feature that would be very useful for extraction of behavioral patterns [18]. Also all tools that were presented in previous section do not record the level of user activity, while this parameter can be effectively used for building user profile [18]. Existing tools that exhibit the categories of accessed Web sites are implemented using client-server architecture, thus private browsing data is reported from user's machine to remote server. That fact may prevent many users from usage of such tools, since not everybody is ready to share confidential data about his browsing preferences. Last, but not least, we are trying to develop a visualization approach that will make the recall and revisit process of Web pages easier, exploiting the mnemonic benefits of patterns memorizing, by showing visited pages as continues chronological graph and not just collection of entries.

Chapter 3

HistoryLane Visualization Principles

This chapter contains the general theoretical background, such as what are benefits of visualization, what should be visualized and also how do we plan to visualize it.

3.1 Theoretical Background

According to Card et al. [1] there are six ways in which visualization can amplify cognition:

1. By increasing the memory and processing resources available to the users: e.g. visualization can be used for storing massive amount of information in easily accessible form, like maps that present geographical information.

2. By reducing the search for information: information can be grouped based on usage models making search faster and more convenient.

3. By using visual representation to improve the detection of patterns: visual organization of data by structural relationships, like time, enhances patterns recognition.

4. By enabling perceptual inference operations: visual representation makes some problems trivial and obvious.

5. By using perceptual attention mechanisms for monitoring: visualization makes possible simultaneous monitoring of multiple events.

6. By encoding information in a manipulable medium: dynamic visualization allows to user select deferent views to highlight particular parameters.

Thus in order to become useful our new visualization approach should address as much as possible of those six visualization concepts. To focus on particular Web browser visualization issue we used list of missing Web browser features that was formulated by Nielsen [23]. Among them we have picked those that are related to history mechanism and are still not implemented by modern browsers or their implementation is not successful:

- Annotations: user should have an ability to add an annotation to previously visited Web pages
- Overview maps: user should be able to inspect his browsing path using graphical structure
- Integrated Search and Browsing: users should be able to define sophisticated queries of history data and receive the result not only in textual format, but also in a graphical one.
- Time-dependent notation: based on appearance of Web page in history it should be easy to find the time spent on the Web site and when it was open.

Another visualization aspect that our novel approach should support is a parallel Web browsing paradigm. According to Huang and White [10] parallel browsing behavior becomes increasingly common among the users, but it is not depicted by modern browsers.

3.2 Visualization Ideas

3.2.1 Basic Shapes

We have chosen to represent single browser window, single tab and single page as rectangles. Since windows may include multiple tabs and single tab may consist of multiple pages, we are going to create a visual hierarchy where the biggest rectangles are windows, they contain smaller rectangles that are tabs, which also contain rectangles representing pages. The visualization is going to be 2D, where axis X is used for time dimension, and axis Y depicts entities (pages,tabs,windows), that stayed open simultaneously (Figure 1.1 presents visualization approach that is close to it). Thus, the more time the window/tab/page stayed open, the wider its rectangle is drawn on the dashboard. The height of tab and page rectangles is constant, so final diagram looks like collection of lanes or strips. That is why we decided to call our approach HistoryLane. The advantage of orienting history diagram on horizontal axis from left to right is that according to Milic-Frayling et al. [21] users have a liner representation of browsing process as a cognitive model in their head, that was mainly formed by "Back" and "Forward" buttons in popular Web browsers. Since these buttons have horizontal arrows pointing to the left and to the right, users intuitively imagine browsing process as a horizontal movement, where pages are placed from left to right in the chronological order.

3.2.2 Mnemonic Hints

We are trying to make recognition of visited Web pages easier. One of most effective ways to do so is to use thumbnail images. According to research performed

by Kaasten et al. [11] thumbnails of Web pages proved to be an accurate way for users to recognize both visited Web site and identify the particular page. For example when users used text only based history list, ratio of successfully recognized Web pages was 60%, but when thumbnails were added users were able to recognize 80% of Web pages correctly.

Despite high efficiency of thumbnails comparing to textual information, we should not neglect provision of page titles and url, because according to Kaasten et al. [11] they also have reasonably high recognition ratio. Titles and urls have almost similar recognition coefficient, thus it is not necessary to provide both of them. That is why we decided to use only page titles, since they are traditionally shorter and help us to save valuable space.

In order to make visual recognition even easier we decided to add a favicon image to page rectangle. Favicons are used in browsers for more than decade and favicon is an important factor in Web site recognition, e.g. Google tried more than 300 permutations of favicons, before they have chose the current one [19].

3.2.3 Information Coding

Numerous studies have shown, that color is the most effective graphical device for reducing visual search time [8]. That is why we decided to use colors for information coding in our visualization approach. As it was mentioned previously in Section 3.1 visualization may be very useful for grouping information. We decided to group Web sites into different categories (news, entertainment, etc.) and code these categories using different colors. User will be able just by single glance identify to what category Web site belongs.

Another usage of colors encoding that we apply is to represent different periods of user activity using different tones of the same color. Let us assume user visited Web page www.bbc.com for 30 seconds. Out of these 30 seconds for first 10 seconds he was active using scrolling wheel, next 10 seconds this tab was in focus, and for the last 30 seconds he switched to another tab. In that case first third of page rectangle will have the strongest tone, the middle one medium tone and the last will be depicted in the palest tone of the same color.

3.2.4 Interaction With a User

A good visualization is not just a static picture that we can walk through and inspect like a museum full of paintings, a good visualization is something that allows us to drill down and find more data about interesting objects [31]. That is why we decided to add features that allow user to manipulate presented data. First one is a zoom in/zoom out feature. Using this feature user can manipulate the size of shown objects and amount of information that is presented, because the bigger rectangle is, more space it has for data presentation.

Another aspect, that is controlled by user, is list of parameters to be shown. User may be interested to see what was his activity period at specific Web page or he may be not. By configuring list of shown parameters user adjusts the system to his own needs making diagram more useful for his custom needs.

Chapter 4

Prototype Implementation

This chapter describes the creation process of prototype which implemented the visualization principles formulated in the previous chapter.

4.1 Requirements

4.1.1 Functional Requirements

Functional requirements dictate the appearance and functionality of HistoryLane prototype. In order to make their perception clearer we grouped them in few subsections, according to their functionality.

General Screen Appearance

1. General screen should be divided into 3 sections: table of history entries as a vertical list, main dashboard with a graphical representation of visited Web pages and a detailed tab analysis area, where single tab can be transferred to from a main dashboard for a detailed analysis.

2. Items in all those sections should be bound to each other, e.g. on moving mouse to page entry in history list, relevant graphical entity should be highlighted.

List of History Entries

1. Visited Web pages should appear in reverse chronological order from up to down.

2. Each entry in the list contains a title text of visited Web page and can be used as hyperlink.

Main Dashboard

1. Main dashboard should contain visualization of windows, tabs and pages.

2. Main dashboard should contain vertical time lines for easy calculation of how much time was spent at particular site.

3. Using the mouse wheel user can expand/shrink the whole diagram.

Time Intervals

1. Browsing history should be presented based on chosen time interval: last hour/day/week.

2. Size of graphical representation should vary according to time interval, the smaller time interval is, the bigger entities are drawn, to exploit the width of the screen.

Windows, Tabs and Pages Representation

1. All objects (windows, tabs, pages) are drawn in chronological order from left to right.

2. Windows, tabs and pages are drawn as rectangles. Window rectangles contains tab rectangles, tab rectangles contain page rectangles.

3. Width of rectangle represents time, that this window/tab/page was open.

4. Color of rectangle indicates the category of Web site (news/social network/e-commerce/etc.).

5. Page rectangle contains the favicon and page title.

6. When mouse is brought to a page rectangle, the thumbnail with a snapshot of page content is shown.

Detailed Tab Analysis

1. Tab should be able to be moved to "tab analysis" area for further inspection.

2. User should be able to add comment to any page entity. After that exclamation mark is shown on this page.

3. Following parameters should be shown about every page: active time, focus time, visit time.

Search Mechanism

1. System should provide to user an ability to search for visited pages based on following parameters: url, title, active time, focus time, visit time. Example of such query is " url contains 'google' ".

2. Entities, that fit search criteria should be highlighted both in list of history entires and visual main dashboard.

4.1.2 Nonfunctional Requirements

Following set of nonfunctional requirements for HistoryLane prototype was formulated in order to make its usage and evaluation easier and more convenient.

1. HistoryLane prototype is implemented as an add-on to existing popular Web browser and not as stand alone application.

2. No data is reported from local machine where HistoryLane is installed to any external entity.

3. HistoryLane prototype is compatible with any of following operating systems: Windows/Mac OS X/Linux.

4. User's browsing process is not influenced by installation of HistoryLane prototype.

5. HistoryLane usage doesn't require installation of any third-party software (graphical libraries, interpreters,etc.).

4.2 Design

Based on functional and nonfunctional requirements we got the decision to develop prototype of HistoryLane system as an add-on for Mozilla Firefox. The main reason for that decision was an availability of browser source code,big amount of existing training materials about how to write such add-ons for Firefox and high popularity of Firefox browser among users.

For creation of visual entities and manipulating them we decided to use HTML 5 technology because of flexibility that is provides and reach API. In order to make code simpler javascript libraries like jQuery and Kinetic were also used. Since Firefox browser doesn't store log data for all events, that we were interested to depict (tab opening time, level of activity at particular Web site, etc.), HistroyLane system listens to all events of Firefox browser and registers them into extended Firefox SQLite database.

4.3 Implementation

4.3.1 Data Collection Mechanism

By default Firefox browser stores browsing data to SQLite data base located at user's profile folder, which is located at C:\Documents and Settings\<username >\Application Data\Mozilla\Firefox\Profiles\<profile folder>\places.sqlite for Windows or /Users/<user>/Library/Application Support/Firefox/Profiles/default.lov/places.sqlite for Mac OS. This data base includes all the data that is required for managing bookmarks and history mechanisms in browser. Since we are interested particularly in history mechanism, let us present what tables store history data :

- **moz_places**: This is the main table of URIs, so every time new URI is opened, new row is added to this table. It contains following fields:
  - *id*
  - *url*
  - *title*: page title
  - *rev_host*
  - *visit_count*: number of visits of this URI
  - *hidden*
  - *typed*: flag either this URI was typed
  - *favicon*: url to favicon image
  - *last_visit*: time stamp of last visit
- **moz_historyvisits**: Table with data about every single visit, thus new row is generated and added to this table every time browser opens any URI. This table contains following fields:
  - *id*
  - *from_visit*: previous page pointer
  - *place_id*: foreign key to moz_places table
  - *visit_date*: visit time stamp
  - *visit_type*: how this visit was done (hyperlink, typed, etc.)
  - *session*: identification of window

Since according to our requirements additional information (like tab opening/closing time) has to be stored, we added few custom tables to this standard data base. They are automatically created during installation of HistoryLane. Here is schema of these tables:

- **windows**: This table contains detailed information about open windows. Here is list of its fields:
  - *window_id*: unique window identifier
  - *open_time*: timestamp of window open time
  - *close_time*: timestamp of window close time
  - *mas_simult_tabs*: maximal number of tabs that were open simultaneously in this window. This data is required to allocate enough space for visualization in the main dashboard.
- **tabs**: Table with log data about tab activities. Here is list of its fields:
  - *tab_id*: tab identifier
  - *window_id*: foreign key from windows table
  - *open_time*: timestamp of tab open time
  - *close_time*: timestamp of tab close time
- **pages**: Table with log data about page activities. Here is list of its fields:
  - *page_id*: page identifier
  - *tab_id*: foreign key from tabs table
  - *window_id*: foreign key from windows table
  - *open_time*: timestamp of page open time
  - *close_time*: timestamp of page close time
  - *visit_type*: specifies how this page was open (typed, new tab, hyperlink, etc.)
  - *active_time*: how much time user was active on this page
  - *focus_time*: how much time this page was in focus in browser
  - *url*: url of page
  - *title*: title of page
  - *favicon_path*: url of favicon
  - *comment*: text of comment added to this page using HistoryLane tool

4.3.2 Showing History

Once the data about Web browser's events is collected and registered, it is ready to be shown at any moment on user's request. After installation of HistoryLane add-on, custom tool bar is added to Firefox browser, click on "Show History" button opens general history screen a new window.

Structure of general screen

General screen contains 3 sections (see Figure 4.1): list of visited Web pages on the left, main visual history dashboard on the right and tab analysis area in a bottom part of screen.

List of Visited Pages

List of visited Web pages contains the visited pages in a reverse chronological order up to bottom. Entries in the list are represented by page titles.

Main Dashboard

As it was mentioned previously, in order to create visual figures we decided to use new features of HTML 5. Since standard functionality of HTML 5 allows only to draw primitive shapes like rectangle, line and text, we decided to use one of sophisticated graphical libraries that are based on HTML 5 technology. We have chosen Kinetic.js, as a library with most flexible API.

Every object (window, tab or page) is drawn as a rectangle and the width of rectangle represents the time it was kept open (geometrical width being directly proportional to a chronological duration), these rectangles are placed in a chronological order from left to right at the main dashboard. Windows rectangles have palest background color and contain multiple tab rectangles. Tab rectangles have darker background color and contain multiple pages. The height of tab rectangle is standard, while height of window rectangle depends on the maximal amount of simultaneously opened tabs. Every page is represented by rectangle of different color depending on its category (search engines - olive, social networks - blue, etc.). Page rectangle also contains a favicon image and a page title (see Figure 4.2). For evaluating how long window/tab/page was open there are vertical time lines with indexation of time intervals.

When mouse pointer is brought into borders of page rectangle, then this rectangle is highlighted, corresponding page entry in list is also highlighted and thumbnail image of Web page screenshot appears, see Figure 4.3.

Detailed Tab Analysis

When user double clicks on tab rectangle, this rectangle is automatically moved to tab analysis area, leaving empty space marked by slightly red color. The main function of tab analysis area is to present particular tab in more detailed manner by allocating to it more space than in the main dashboard and by provision additional information about that tab. When tab is moved to tab analysis area, level of user activity during page visit is depicted by rectangles with different undertones of the same color. The hierarchy is: active time ->focus time->visit

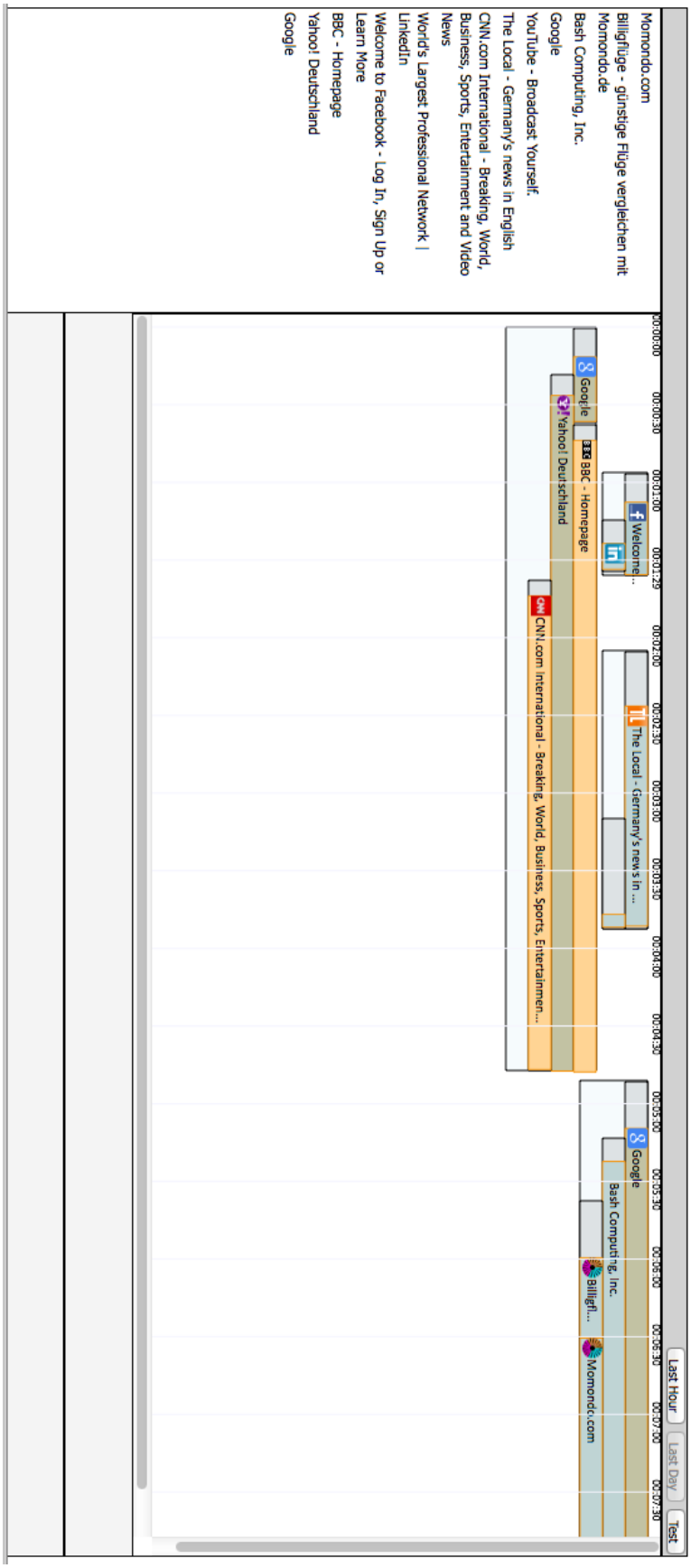


Figure 4.1: A general screen of HistoryLane tool

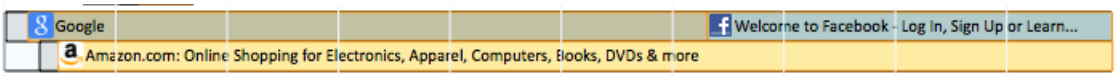


Figure 4.2: An example of single window representation. This window contains two tabs. Upper tab contains two pages

Figure 4.3: On page selection

time, active time is represented by darkest undertone, visit time by palest one, see Figure 4.4.

Another feature that is available to user, when he uses a tab analysis area, is an option to show numerical values of active time/focus time/visit time or to add a custom comment. To do so, user clicks with a right click on desired page rectangle, after that a new pop up menu is shown to him, see Figure 4.5.

After that pop-up menu is closed, the selected parameters are shown at page rectangle accompanied by an exclamation mark, see Figure 4.6.

Search Capabilities

To implement "Search" feature first we defined list of parameters, that are going to be used in filtering conditions. We decided to provide the user with an ability to define search queries based on page url, title, active time, focus time and visit time. In addition any combination of these parameters may be used. For example search query *"active_time >36"* returns all pages where user was active for more than 36 seconds, query *"visit_time <20 url contains 'google' "* returns all pages that belong to google domain and their visit time was less than 20 seconds. We implemented simple interpreter for parsing query text and translating it to xPath conditions (since log data is stored in XML format). Finally search results are shown both in chronological list of history pages and in the main dashboard by using blue highlighting, see Figure 4.7.

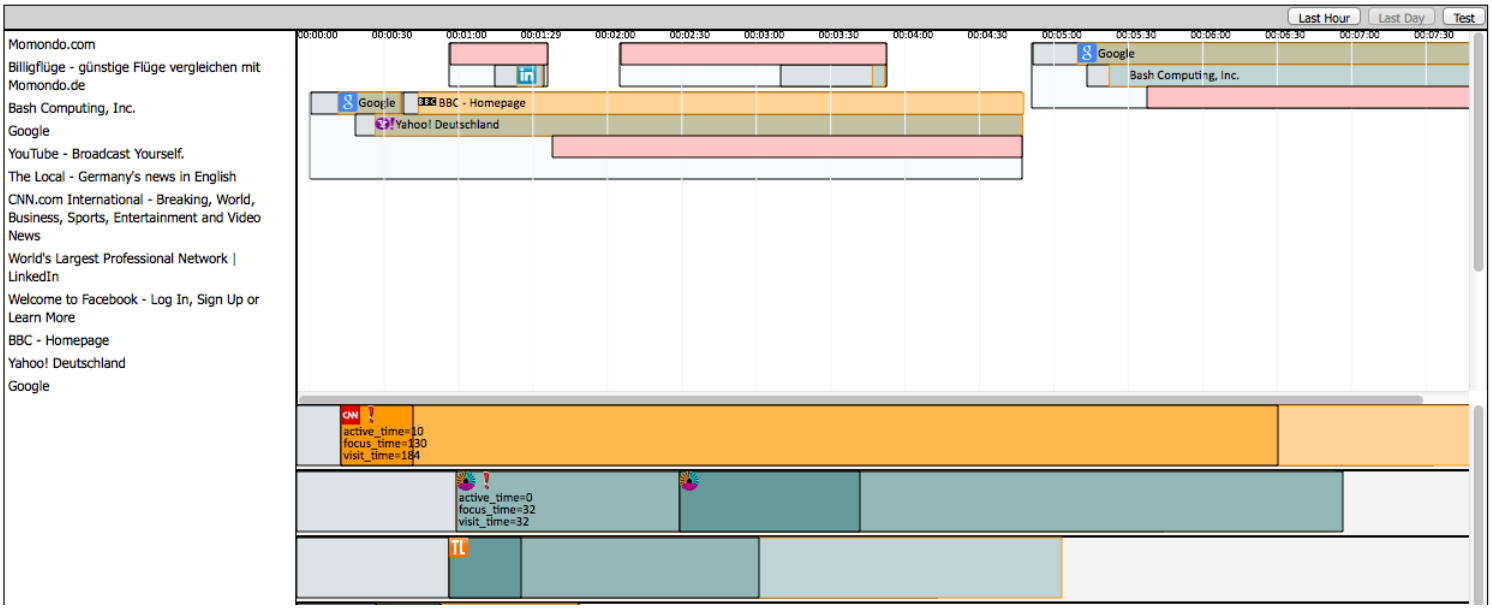


Figure 4.4: Single tab can be moved to tab analysis area. In that case, more detailed visualization is provided. Missing spot in main dashboard is filled with red rectangle.

4.4 Testing

Before we start to describe the testing activities, we have to mention that for development of the prototype we used an iterative lifecycle. Iterative lifecycle implies that set of requirements is not cast in stone, but may change during product development. Some of our visualization ideas looked good on paper, but did not live up to expectation upon implementation, so we had to adjust the requirements according to the intermediate results.

Every iteration had four main phases: Requirements, Design, Implementation and Review. Most of the testing activities were performed during Implementation and Review phases.

During Implementation phase we performed following test actives:

- Unit testing: to test new added features.
- Integration testing: to test how new features work, when they work together with previously existing components.
- System testing: to test complex user oriented scenarios, that include usage of the new feature. An emphasis was put on performance testing.

During Review phase we eliminated the unit testing and concentrated on integration and systematic testing:

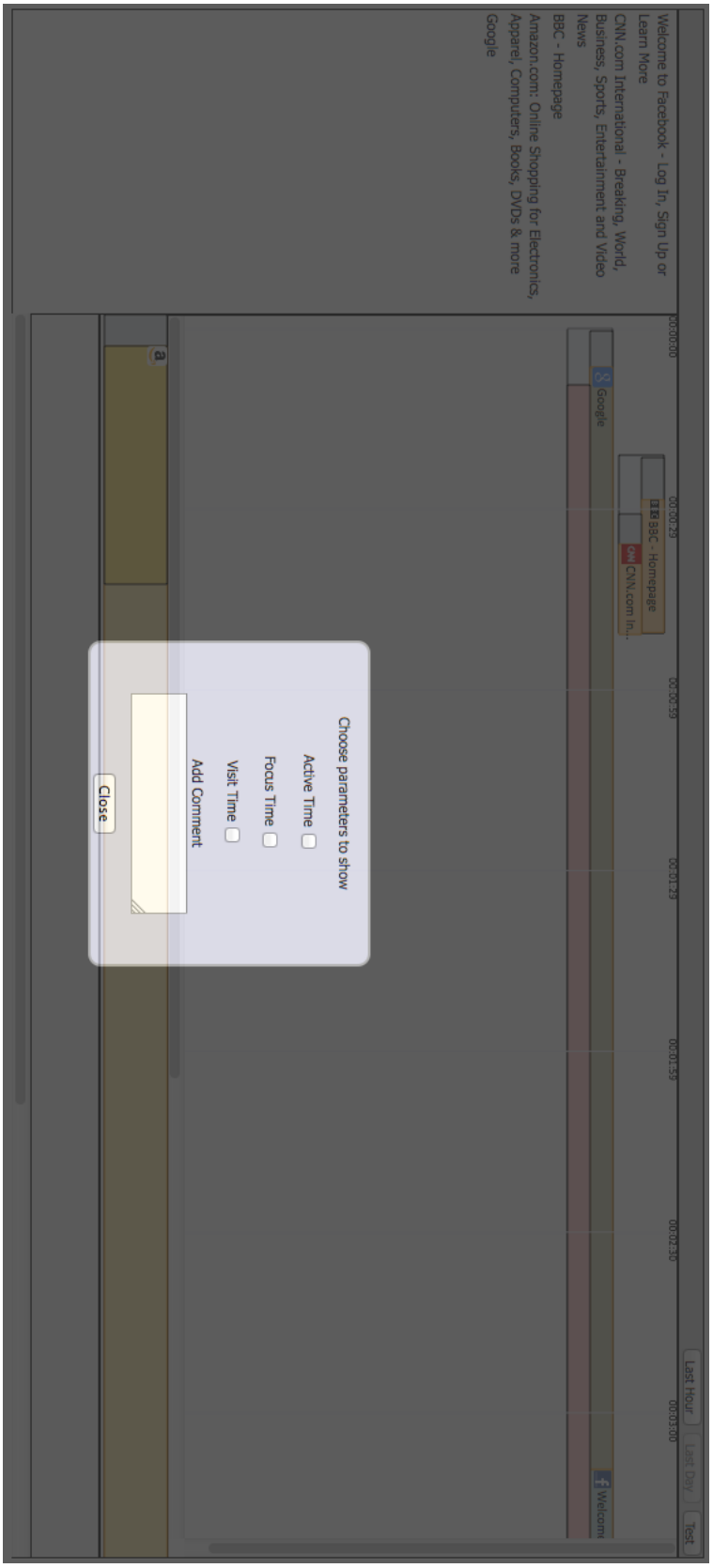


Figure 4.5: Pop up for defining shown parameters of a single page in tab analysis area.

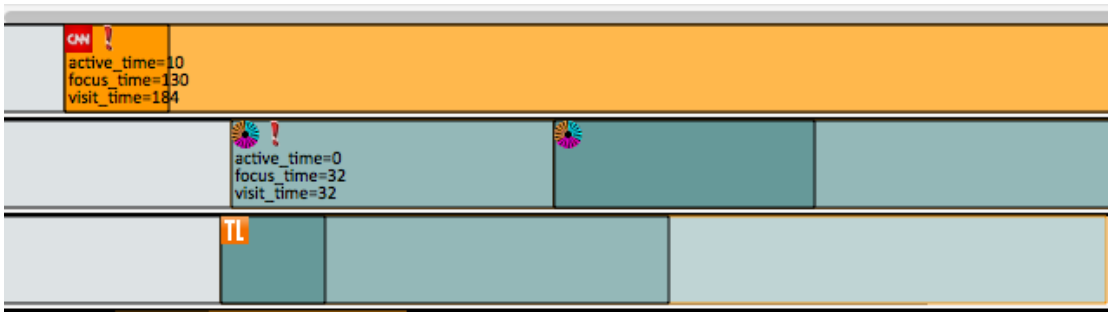


Figure 4.6: Selected parameters are added to a page rectangle.

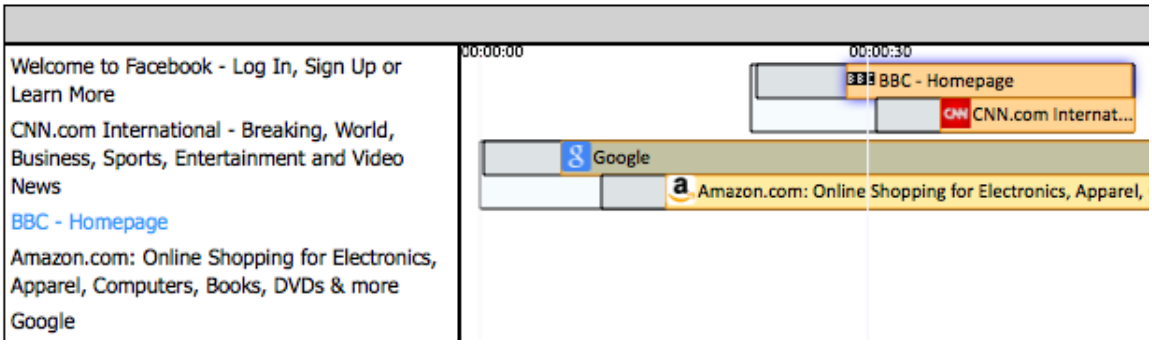


Figure 4.7: Exhibition of search results.

- Integration testing: to test how previously developed modules work in co-operation with a new one.
- System testing: to test complex user oriented scenarios, that include usage of new feature. Particularly we were looking for side effect bugs.

When the prototype was complete, we also tested the installation and deployment scenarios to minimize the risk that experiment participants will have any problem with installing the prototype on their machines. Installation scenarios both for Mac OS and Windows 7 were executed.

Chapter 5

Research Methods

This chapter describes what research methods did we use in order to answer the research questions and validate our novel history visualization approach.

5.1 Introduction

In Chapter 1, describing the scope of this master thesis we stated that we are interested to formulate new principles of browsing history visualization, to build the prototype and to evaluate these principles using a working prototype. Applying the research methods that are presented in this chapter we try to get a good understanding about quality and successfulness of our visualization approach. We provide the detailed description of the research strategy adopted to find answers for the research questions, including how data is collected and analyzed. In addition we present potential limitations and problems of chosen research strategy and its implementation.

5.2 Research Strategy

According to Kothari [14] research strategy should be formed in respect of following:

- The means of obtaining information
- The availability and the skills of the researcher
- The objective of the problem to be studied
- The availability of time and money for the research work

We analyzed these aspects and formulated following findings:

- Main means of obtaining information are exploration of literal resources and collection of inputs from Web browsers users.

- This research is going to be conducted by master student under guidance and supervision of professional researches. The master student has basic research skills and may need consultation of supervisors regarding research structure design and results analysis.
- This is an exploratory research study, wherein the major emphasis is on discovery of new ideas.
- Time period allocated for this master thesis is six months and there is no money involved.

In general research designs can be divided into 3 categories: 1) research design in case of exploratory research studies; 2) research design in case of descriptive and diagnostic research studies; 3) research design in case of hypothesis-testing research studies [14]. For this thesis we used research strategy that combined elements both of exploratory research and hypothesis-testing.

First we performed structured literature review to identify the exact problem and to find a gap that should be filled by this thesis. Then we devised visualization ideas and proposed a hypothesis, that HistoryLane visualization approach is better than existing alternatives. For validating this hypothesis we created a working prototype and made it available to users. Next step was to validate the hypothesis using appropriate tools, we decided to do it using survey and experiment.

5.3 Survey

5.3.1 Design

We have chosen a survey, because it has several claimed attractions [3]:

- it is economical and efficient
- generates numerical data which can be processed statistically
- gathers standard information
- supports or refutes hypotheses about the target population
- generates accurate instruments

Using survey we are seeking to gather large-scale accurate data from as representative sample as possible in order to say with a measure of statistical confidence about opinion of target group about certain statements. Our survey is a confirmatory survey, since it is designed to confirm our hypothesis.

There are three factors that should be taken into account for a survey design: specification of the exact purpose of the inquiry, population and resources available [3].

The exact purpose of our survey would be formulated as *"to obtain opinion of potential users about HistoryLane visualization approach according to defined parameters"*. Since Web browsers are widely used by general population we had no need to form particular group to which the inquiry is addressed. As it was already mentioned in this chapter there is no budget for the thesis, so survey had to be created using free available tools and participants are not paid for their contribution.

It makes sense to use an Internet survey, because Internet surveys have following claimed advantages [3]:

- They are cheap
- Less time required for distribution, gathering and processing data
- Respondents can complete questionary from home
- There are tools for making surveys attractive (e.g. graphics)

There is a set of recommendations for Internet based survey, formulated by Cohen et al. [3] that we followed when created our survey. Survey should not contain too many open-ended questions where users have to type their answer. Also the introduction to the questionnaire should be short (no more than one screen) and informative, without too many instructions. The very first question tends to raise in respondent's mind a particular mind-set, that is why it should be formulated very thoroughly. Web questionnaire should be started with a welcome screen that motivates the respondents to continue. There should be no differences in the visual appearance of questions. The sequence of questions should be logical and continuous.

5.3.2 Data Collection

As a technical platform for creation of the Internet survey we used Google Docs Form, because of its simplicity and flexibility. It allows to moderate survey easily, so first we created a pilot version, that was later slightly changed to improve the usability. In addition, for making a task more clear, it included video file with the exact instructions. Full text of survey can be found in Appendix section. First respondents watched the video tutorial about how to install and use the prototype, then they were asked to install it on their local machines. Next step participants had to use Firefox for active browsing for 45 minutes in order to create history of browser events. Later they used HistoryLane prototype to observe their browsing history. As a last step respondents filled up questionnaire to express their opinion about it. This process is shown on Figure 5.1.

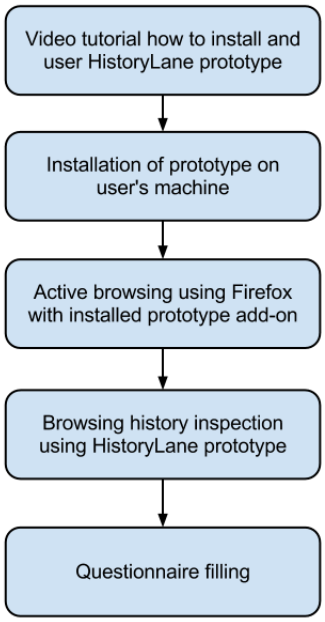


Figure 5.1: Survey data collection process

Sampling universe for our survey theoretically is a whole population of Earth and sampling unit contains two billions of active Internet users. However we had problems with attracting responders, that is why our sample size was limited to 17 participants. To keep things simple we used volunteer sampling, trying to create maximal diversity of participants in geographical, gender and educational perspectives.

5.3.3 Framework for Data Analysis

Data Editing

Before undertaking any data analysis, responses should be checked for consistency and completeness [13]. So first we check survey questionnaires, looking for incomplete or inconsistent answers, these questionnaires are omitted.

Data Classification

In the survey we have both questions where participants provide numerical scores and open question where they write their opinion as a plain text. These two categories are going to be analyzed separately, since for the first one we are going to use statistical tools, whereas the second one is going to be analyzed through informal reading.

Data Analysis

According to [13] it is a common practice to convert the ordinal survey scale to numerical equivalent and to analyze the data as set of numerical values, however this approach violates the mathematical rules for analyzing ordinal data. That is why we decided to present results of survey as the proportion of responses in each category. This is the most informative and intuitive way to present raw data. In addition we calculate mean and variance values for every metric.

5.3.4 Validity Threats

Internal Validity Threats

We decided to learn users' opinion about HistoryLane visualization approach using an Internet survey, while by interviewing them theoretically we could collect more detailed data. However due to time limits we could not interview more than 10 users, while by using the Internet survey we collected data from more respondents.

Internet based surveys are not perfect and have their problems [3]:

- Respondents may send multiple copies of answer
- The order of items affects respondent's answer
- Respondents may not understand instructions correctly
- Respondents may misunderstand specific questions

We tried to address all these issues by sticking to guidance for Internet surveys.

In addition users provided their opinion about HistoryLane after relatively short usage period, so their opinion would be different, if they were using it more time.

External Validity Threats

We are aware that since Web browsers are extremely popular and are used by billions of users, sample of few tens of participants is not representative enough. In addition volunteer sampling policy compromises the generalizability or representativeness of research [3]. Most of those volunteers come from same socio-economic background and have approximately same education level. This makes generalization of results problematic.

5.4 Quantitative Experiment

5.4.1 Design

Experiments can be classified into two broad categories based on experimental design: informal experimental design and formal experimental design. Informal experimental designs are those designs that usually use less sophisticated form of analysis, whereas formal experimental designs offer more control and use rigorous statistical procedures for analysis [14]. Taking into account, that we are lacking experience in research methodology, we decided to use informal experimental design because of its simplicity.

In order to evaluate either HistoryLane visualization approach is effective or not, we compare HistoryLane with other alternative visualization approach which provides similar functionality. Since we try to compare two methods it makes sense to use paired comparison design, because a pair of treatments are compared in each single task.

A paired comparison experiment design is a randomized block design with blocks size of two. Within each block two treatments are randomly assigned to two test units [33].

5.4.2 Data Collection

First, we formed two groups: control group and test group, 7 participants in each one. Then participants in the test group received instruction to install HistoryLane prototype on their computers and watched short tutorial about how to use it, whereas participants in control group received guidance how to install an alternative history tool, Google Chrome HistoryStat 1.3 extension. In addition users of control group were allowed to use other browsing history mechanism. HistoryStat was selected, because it provides some features that are parallel to features of HistoryLane. After that both control and test groups participants had to complete quest which contains multiple tasks that have to be performed using Web browser. An example of such task is *"Find in what european city is located the bridge that is shown on the image"*, these tasks are designed to be solved not by a single search engine request, but by examining at least few Web sites. This is done in order to generate browsing history that simulates history created as result of normal browsing process based on some user's goals. A fragment of such history generated during the quest by one of the participants using HistoryLane is shown in Figure 5.3. After that users created browsing history, the core experiment starts. Users of both groups had to answer questions measuring time that is required for that. Typical question is *"In what Web page did you complete the task #3 from the quest ?"*. According to our hypothesis HistoryLane prototype presents data in the way that improves users perception and makes history analysis easier and faster. Thus dependent variable in our

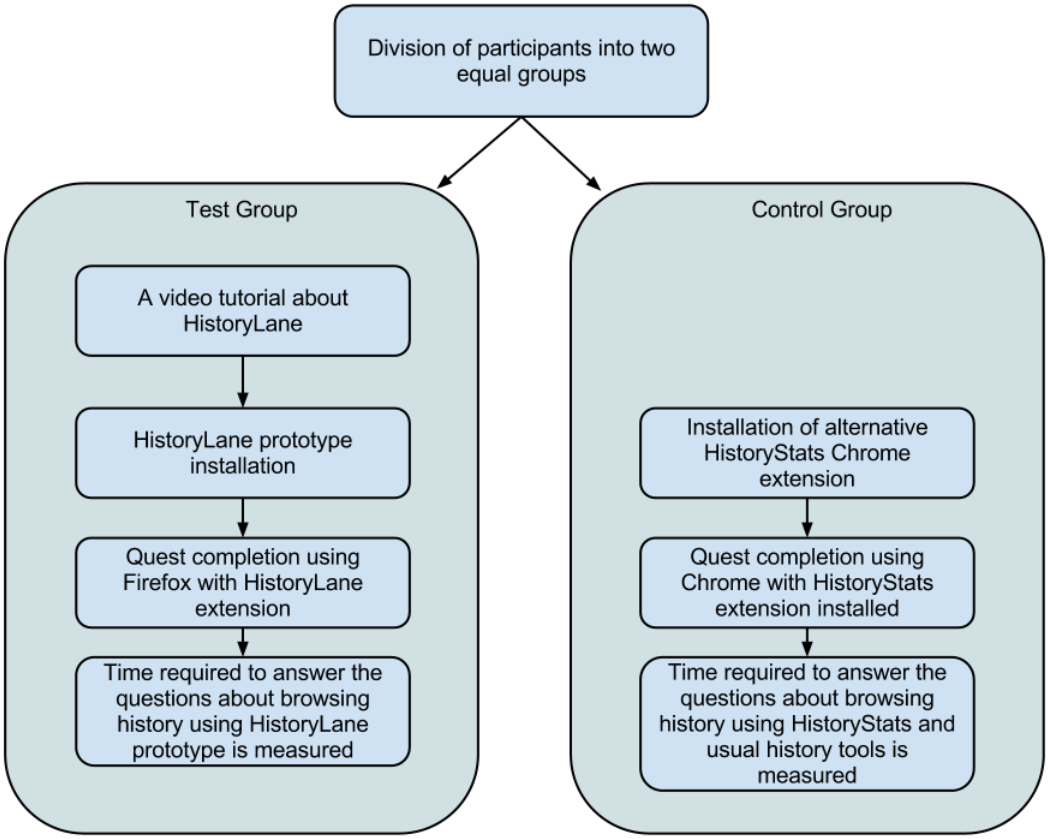


Figure 5.2: Data collection schema

experiment is a time required to find an answer for a question. Figure 5.2 presents the schema of the experiment execution, the full text of the quest and experiment questions are available in Appendix to this thesis.

5.4.3 Framework for Data Analysis

Data Editing

Experiment participants report experiment results by themselves and their answers may be provided in different formats. We clean that data and record reported time in seconds using single standard.

Data Classification

We have two classes of data: time metrics reported by users of HistoryLane prototype and time metrics provided by users that used HistoryStats Google Chrome extension accompanied by traditional history tools.

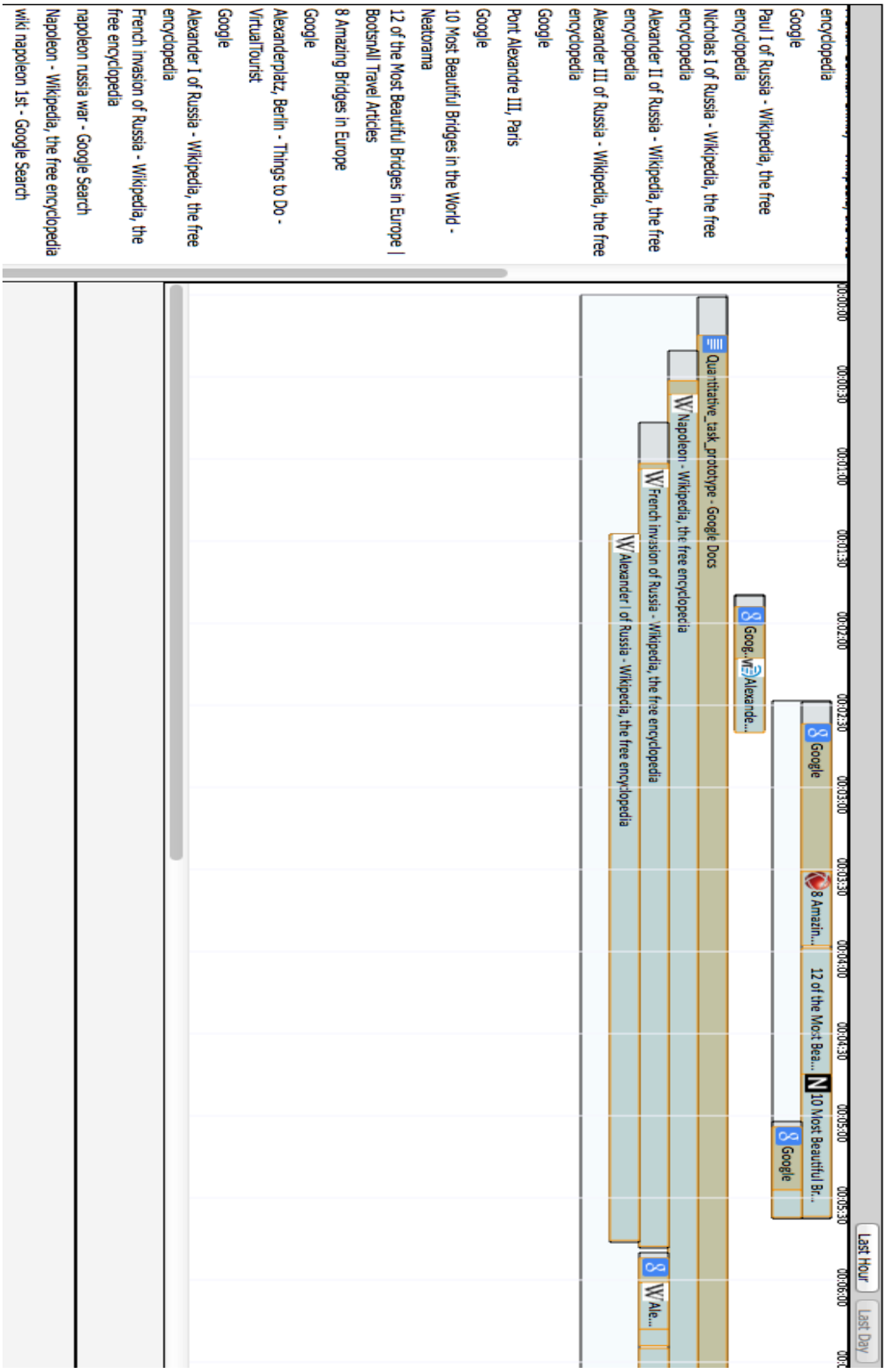


Figure 5.3: Fragment of history generated during quantitative experiment

Data Tabulation

Tabulation is a procedure of arranging data in a concise and logical order [14]. In our case tabulation is done automatically using Google Forms platform. All raw data is presented in two tables: one with results received by usage of HistoryLane prototype and the second one with results that were reported by users that used an alternative visualization tool.

Data Analysis

Two samples are considered to be independent, when participants are randomly assigned to groups, without regards who else has been selected for either sample [16]. Since we created sampling groups for our experiment absolutely randomly they can be called independent. The most widely used and the most widely known statistical test for independent samples analysis is independent samples t test [17]. We apply this method calculating what is the difference in mean time required to complete each single experiment task using HistoryLane and the alternative tool.

5.4.4 Validity Threats

Internal Validity Threats

Due to logistic reasons users are going to participate in the experiment not in controlled environment, but in their homes. Thus we can not control that they measure time that it takes to answer the question correctly and that they report an accurate time. In addition the performance of users in experiment tasks in influenced by the hardware they possess and the level of their engagement.

Some of participants know each other and could discuss the content of experiment prior to participation, since the experiment was not conducted simultaneously with all participants, but during relatively long period of time.

Participants in the control group (the one that used HistoryStats tool) did not receive any training about its usage, since as a standard Google Chrome extension it should be self explaining, while participants in control group watched the video tutorial. Such uneven approach may also be a threat for experiment's internal validity.

According to feedback that we received from participants, it was hard for them to perceive the structure of the experiment, thus their is a probability that some of participants did not understand tasks completely and their responses contain an inaccurate data.

External Validity Threats

We are trying to prove the hypothesis, that HistoryLane visualization is more effective for analyzing browsing history than other existing history tools. But since we can not compare "apples with oranges" we should compare HistoryLane to similar visualization approaches. We tried to find the closest analog to our system and believe that HistoryStats Google Chrome extension may be such analog. However these two visualization approaches were created using different set of requirements, thus they do not provide exactly the same functionality. In addition the functionality of Mozilla Firefox and Google Chrome differ, e.g. Chrome supports auto completion of search queries directly in url field, which may make search more effective.

For this experiment we also used volunteer sampling, thus opinion of participants may be biased towards the HistoryLane prototype due to personal sympathy. Also the size of both control and test groups was only 7 users due to difficulty to attract participants. That is why results of any of those groups may be strongly influenced by a single participant with extremely splendid or inferior performance. As it was already mentioned in Section 5.3.4 it is not clear to what extent students represent the general population.

Another solid threat to external validity of this experiment is the way we stimulate/simulate the generation of browsing history. We force users to complete certain quest which consists of multiple tasks, but the generated browsing path may not fit the natural behavior of typical user.

Experimental arrangements may also jeopardize the external validity, since participants are aware of experiment conduction and their behavior may differ from normal browsing process.

Chapter 6

Research Results

In this chapter we present the results of HistoryLane visualization method evaluation received trough survey and quantitative experiment.

6.1 Survey Results

Our survey contains 11 statements and two open question where users could express their opinion by providing a textual feedback . The exact text of survey can be found in Appendix B section. All statements had a scale from 1 to 10, where 1 stands for "Completely disagree" and 10 stands for "Completely agree". For simplicity of survey analysis we mapped the research questions to to survey statements, this mapping is presented in Table 6.1.

Table 6.1: Mapping between research and survey questions

Research Question	Survey Question
RQ1: Does the aggregation of browsing history into tab sections makes the perception of this history more convenient for users?	I browse multiple tabs or windows in parallel often?
RQ1: Does the aggregation of browsing history into tab sections makes the perception of this history more convenient for users?	I find that grouping of visited pages into tabs and tabs into windows for visualization is useful/intuitive/insightful.
RQ2: How to visualize collected history data?	Representation of visited pages from left to right in chronological order is useful/intuitive/insightful.
RQ2: How to visualize collected history data?	Accompaniment of visited pages using favicons (those small icons from address bar) helps me a lot to identify the visited Web page.

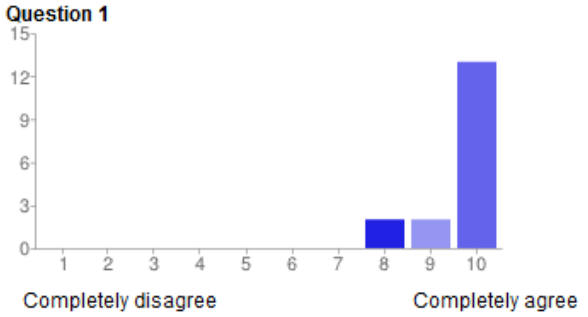
Continued on next page

Table 6.1 – continued from previous page

Research Question	Survey Question
RQ2: How to visualize collected history data?	Accompaniment of visited pages using thumbnails helps me a lot to identify the visited Web page.
RQ2: How to visualize collected history data?	Accompaniment of visited pages using page titles helps me a lot to identify the Web pages.
RQ2: How to visualize collected history data?	Usage of different colors for different categories of Web sites helps me to analyze patterns of my browsing behavior.
RQ2: How to visualize collected history data?	Vertical time lines are useful for me and help me to analyze patterns of my browsing behavior.
RQ2: How to visualize collected history data?	List of visited pages (on the left side) is bound to drawn entities (on the right side) in an intuitive manner
RQ2: How to visualize collected history data?	When I move a tab to "Tab Analysis Zone" (using double click), the representation of active_time/focus_time/visit_time using different gradation of the same color is intuitive to me.
RQ2: How to visualize collected history data?	An ability to expand/shrink visual entities using mouse scrolling and see additional details is useful.
RQ3: What are the benefits of browsing history visualization?	What did you like about this visualization approach?
RQ3: What are the benefits of browsing history visualization?	What would you change or improve in this visualization approach?

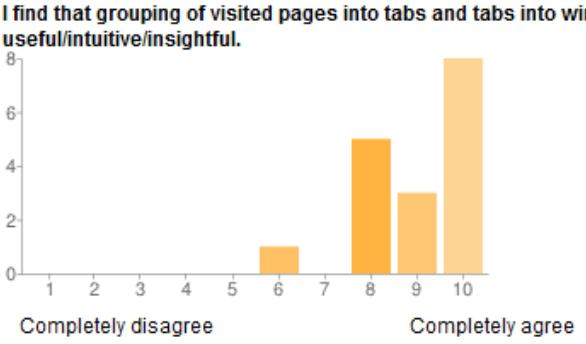
All survey participants agreed that they are browsing multiple tabs or windows in parallel often. Out of them 76% agreed with it completely, see Figure 6.1, reinforcing theory that modern users exploit actively parallel browsing feature of Web browsers.

The core of our visualization approach is a principle of grouping of pages into tabs and tabs into windows, that they belong to. An absolute majority of participants (94%) finds this principle insightful. The responses for statement *"I find that grouping of visited pages into tabs and tabs into windows for visualization is useful/intuitive/insightful"* had average grade of 9 with variance 1.375. The mean value was also 9, the full distribution of responses is shown in Figure 6.2.



1 - Completely disagree	**0**	0%
2	**0**	0%
3	**0**	0%
4	**0**	0%
5	**0**	0%
6	**0**	0%
7	**0**	0%
8	**2**	12%
9	**2**	12%
10 - Completely agree	**13**	76%

Figure 6.1: Responses to statement *"I browse multiple tabs or windows in parallel often"*



1 - Completely disagree	**0**	0%
2	**0**	0%
3	**0**	0%
4	**0**	0%
5	**0**	0%
6	**1**	6%
7	**0**	0%
8	**5**	29%
9	**3**	18%
10 - Completely agree	**8**	47%

Figure 6.2: Responses to statement *"I find that grouping of visited pages into tabs and tabs into windows for visualization is useful/intuitive/insightful"*

Thus based on these findings we can conclude that we have found a good way to aggregate visited Web pages.

Respondents of survey answered that positioning of visual entities in chronological order from left to right is insightful for them. All of them, but one, agreed to that at some extent. The average grade for statement *"Representation of visited pages from left to right in chronological order is useful/intuitive/insightful"* was 8.88 with variance 3.11. The median value was 10, so at least half of participants intuitively understand this approach. The full distribution of responses is shown in Figure 6.3.

Trying to help the users to identify and differentiate the Web pages we put favicons to page rectangles on the diagram. This feature was considered useful by all respondents (all grades were higher than 6), even so some of them find it

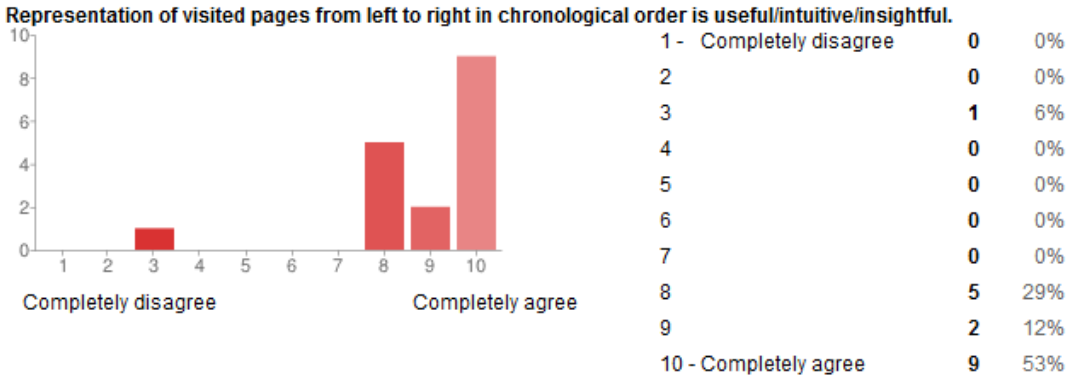


Figure 6.3: Responses to statement *"Representation of visited pages from left to right in chronological order is useful/intuitive/insightful"*

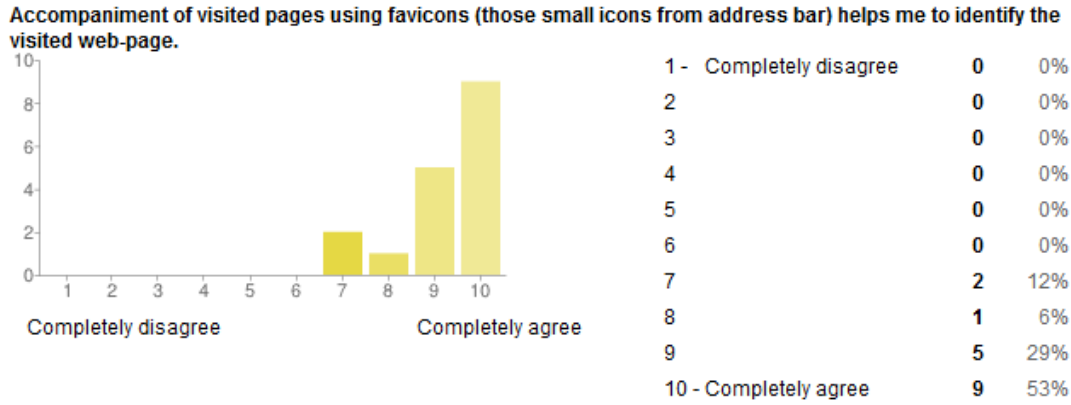
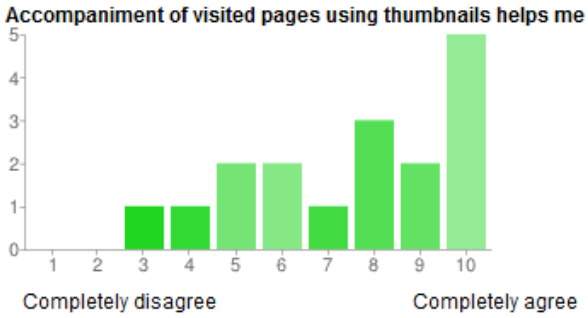


Figure 6.4: Responses to statement *"Accompaniment of visited pages using favicons (those small icons from address bar) helps me to identify the visited Web page"*

partially useful. The average grade given by respondents to *"Accompaniment of visited pages using favicons (those small icons from address bar) helps me a lot to identify the visited Web page"* survey statement was 9.23 with variance 1.06, median value was 10. The exact distribution of responses is shown in Figure 6.4.

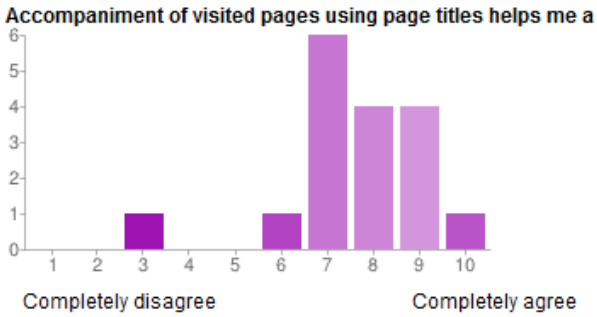
Another way to improve the cognitive recognition of Web pages is the usage of thumbnail image preview, that is shown, when the mouse brought over page rectangle. Surprisingly 18% or respondents do not consider that thumbnail images help them to identify the visited Web pages. Most of users find this feature useful, among them 30% completely agree that thumbnails help them to identify visited Web pages, see Figure 6.5.

Web page title is embedded into page rectangle next to favicon. Most of the respondents answered that it helps them to identify visited Web pages, but



1 - Completely disagree	0	0%
2	0	0%
3	1	6%
4	1	6%
5	2	12%
6	2	12%
7	1	6%
8	3	18%
9	2	12%
10 - Completely agree	5	29%

Figure 6.5: Responses to statement *"Accompaniment of visited pages using thumbnails helps me to identify the visited Web page"*



1 - Completely disagree	0	0%
2	0	0%
3	1	6%
4	0	0%
5	0	0%
6	1	6%
7	6	35%
8	4	24%
9	4	24%
10 - Completely agree	1	6%

Figure 6.6: Responses to statement *"Accompaniment of visited pages using page titles helps me to identify the Web pages"*

the percentage of those who completely agree that page titles significantly help to identify visited Web pages is only 6%. Also 6% of respondents disagree, that titles help them to identify visited Web pages. The exact distribution of responses is shown in Figure 6.6.

To make the identification of Web site category easier we introduced different colors for different categories of Web sites. There was variety of responses about this idea by users. Some of them (6%) completely disagree with this approach, 30% of respondents nether agree nor disagree and 64% agree (approximately half of them completely) that usage of different colors helps to analyze patterns of browsing behavior, see Figure 6.7.

In order help users evaluate how much time they have spent at particular Web page there are vertical time lines, that create grid above the main dashboard. Most of users find these time lines useful, however only 12% of them completely

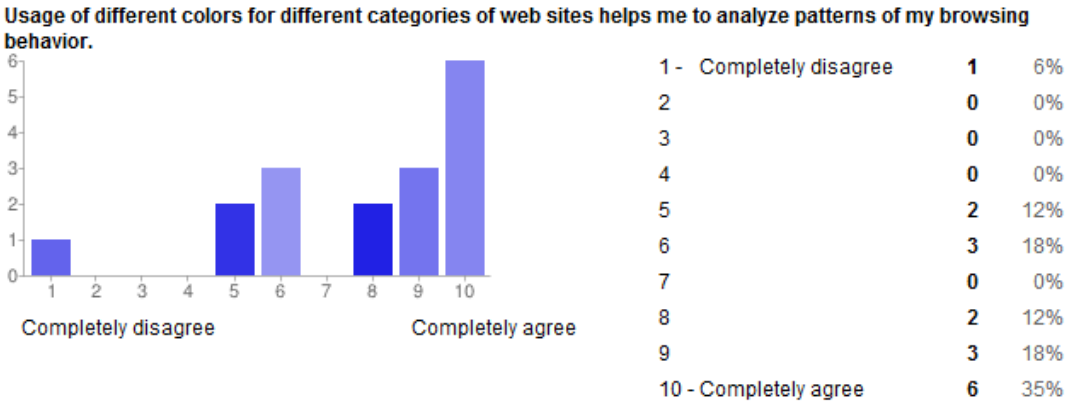


Figure 6.7: Responses to statement *"Usage of different colors for different categories of Web sites helps me to analyze patterns of my browsing behavior"*

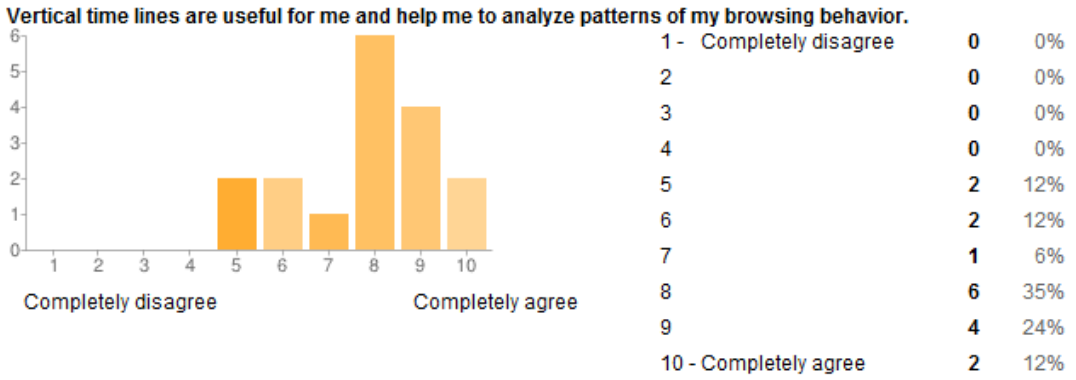


Figure 6.8: Responses to statement *"Vertical time lines are useful for me and help me to analyze patterns of my browsing behavior"*

agree with statement *"Vertical time lines are useful for me and help me to analyze patterns of my browsing behavior"*. The average grade that was given is 7.93 with variance 2.44. The value of median was 8, see Figure 6.8.

Our approach combines traditional way to represent history as a list of entries with a graphical dashboard, where pages are drawn as rectangles. These two mechanisms are bound and once page is selected in one of them, it is also selected in another one. However 36% of users stated, that this is not implemented in an intuitive manner, out of them 18% completely disagree that it was done successfully. Most of respondents gave grades 7 and 8 to statement *"List of visited pages (on the left side) is bound to drawn entities (on the right side) in an intuitive manner"*, see Figure 6.9.

When tab is moved to "detailed tab analysis" area, parameters like active time, focus time and visit time are visualized using different brightness levels of

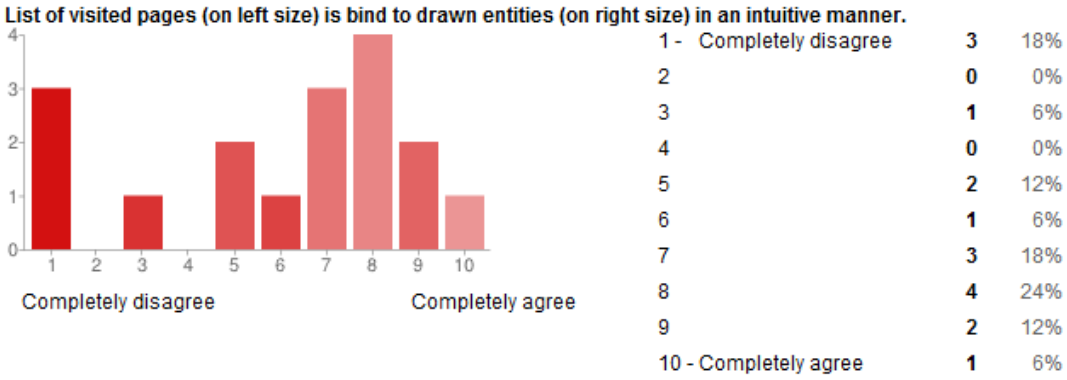


Figure 6.9: Responses to statement *"List of visited pages (on the left side) is bound to drawn entities (on the right side) in an intuitive manner"*

the same color. This approach for visualization received ambiguous grades, the average grade for statement *"When I move a tab to Tab Analysis Zone (using double click), the representation of active_time/focus_time/visit_time using different gradation of the same color is intuitive to me"* was 7.23 with variance 4.06. However most of users find this approach intuitive, since the median value is 7. The exact distribution of responses is shown in Figure 6.10.

Using mouse wheel scrolling users can manipulate the size of visual entities, zooming in and out. This feature is considered useful by 94% of users. The average grade 8.17 was given to this feature, but the variance of grades was extremely high - 5.4. The value of median is 9, which shows that this feature is perceived as highly useful. All users' responses are shown at Figure 6.11.

In the second part of survey respondents were asked to provide their opinion about HistoryLane visualization approach using free text format. Those are the key points of positive responses:

- It is intuitive to use
- Usage of thumbnails is useful
- Extended search functionality is a powerful tool
- The hierarchal representation window>tab>page is intuitive
- HistoryLane is useful for analysis of the time spent on the Web sites
- HistoryLane is good to analyze level of activity
- Easy to identify visited sites
- Usage of time lines is a good idea

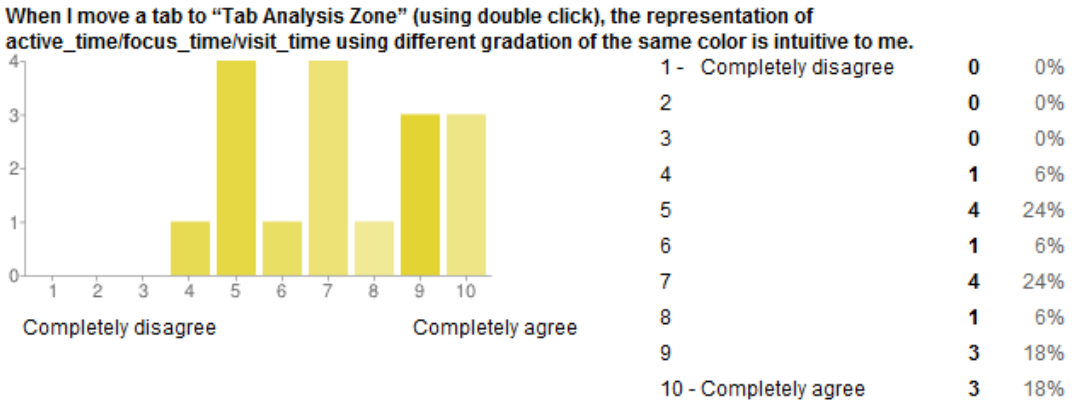


Figure 6.10: Responses to statement *"When I move a tab to Tab Analysis Zone (using double click), the representation of active_time/focus_time/visit_time using different gradation of the same color is intuitive to me"*

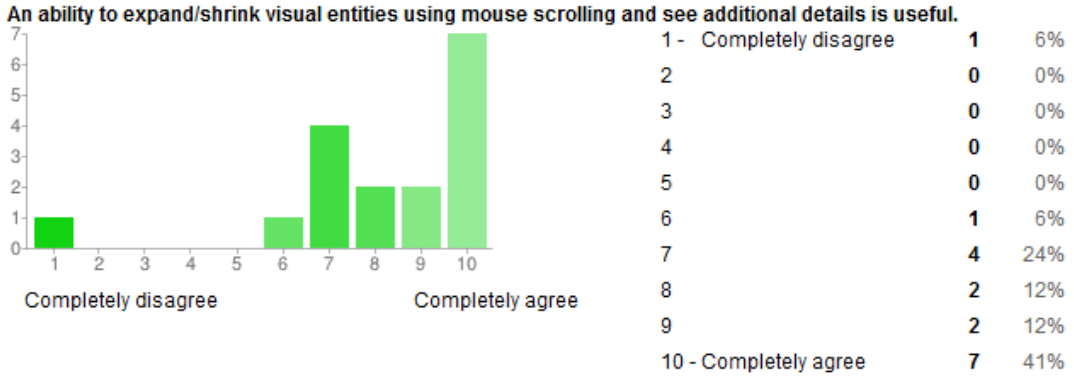


Figure 6.11: Responses to statement *"An ability to expand/shrink visual entities using mouse scrolling and see additional details is useful"*

Beside the positive feedback we also received list of things, that users would change in HistoryLane. Among them following items are present:

- It would be good to see the graph of dependencies between Web pages
- There is no functionality for history manipulation (delete/edit)
- No graphical charts showing browsing trends
- It would be nice to group pages that belong to the same domain
- Background should be configurable
- UI is not self explaining
- It is hard to memorize what color symbolizes what category
- Page entires in a list may be grouped

Results of survey show that majority of respondents find our visualization approach helpful and intuitive. Respondents approved that the representation of history pages, their grouping and positioning are informative and natural. An interesting result was relatively high amount of respondents, that find usage of thumbnails ineffective in identification of visited pages, since authors of this thesis considered it as one of the most useful features of HistoryLane. Another feature that users were not very enthusiastic about, despite expectations, is a usage of different colors for different categories of Web sites. Among things that should be improved according to users' responses, is a connection between the list of history entries and graphical objects in the dashboard, 36% of respondents stated that it is not implemented in an intuitive manner. Another aspect that requires future research is the usage of mouse wheel for zooming in and out, since the variance of grades was extremely high, which means that some users like it very much, while some users didn't like it at all. The answers gathered through open question provide a good material for defining directions of future work.

6.2 Quantitative Experiment Results

In the quantitative experiment we measured how much time does it take to users to perform specific tasks related to Web browsing history using HistoryLane and an alternative tool. The full text of experiment instructions is available in Appendix section. The summary of results is shown in the Table 6.2 and Figure 6.12.

Table 6.2: Summary of quantitative experiment results (in seconds)

	Task#1	Task#2	Task#3	Task#4
HistoryLane	μ=21.1 σ=22.4	μ=42.0 σ=19.7	μ=29.7 σ=17.0	μ=63.1 σ=43.4
Alternative tool	μ=30.3 σ=12.5	μ=47.6 σ=23.2	μ=29.0 σ=11.2	μ=42.6 σ=28.2

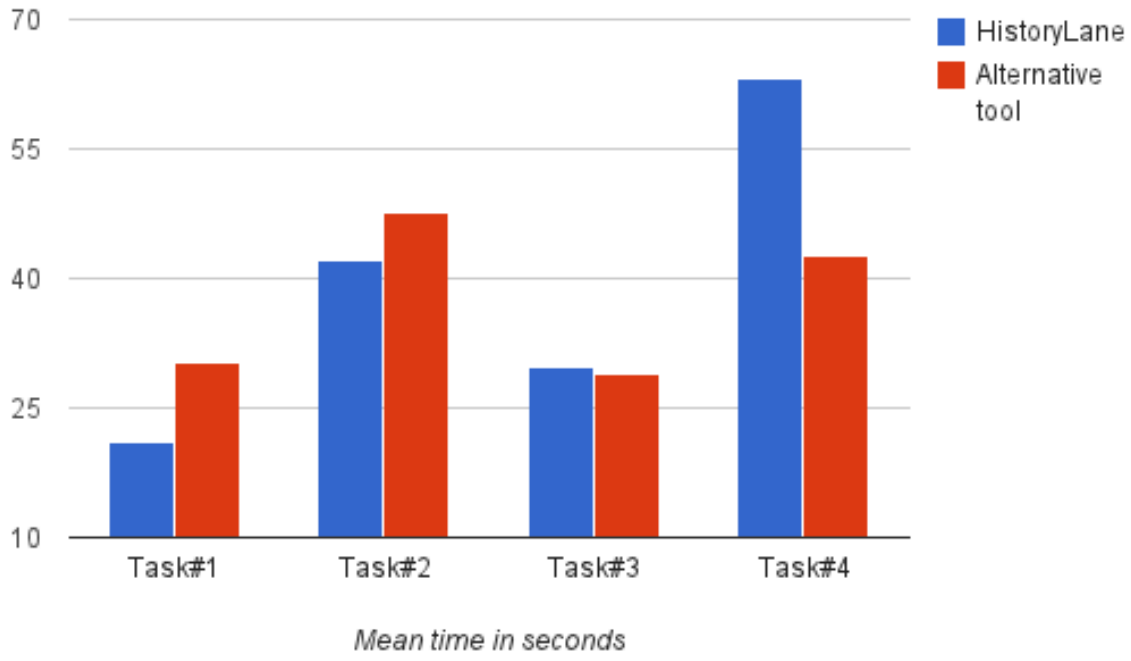


Figure 6.12: Quantitative experiment results chart

First task was to locate the exact page where user has found an answer for question #3 in quest. Using HistoryLane prototype users have accomplished it within mean time μ=21.1 seconds and standard deviation σ=22.4. Using an alternative HistoryStats tool the mean time was μ=30.3 seconds and standard deviation σ=12.5. The results of t test show that with level of confidence p=0.64 we can state, that mean time of this task execution using HistoryLane is shorter than using an alternative tool.

Second task was to determine the amount of times that the answer for a quest challenge was found directly on the first opened page. Using HistoryLane prototype users did it with mean time μ=42.0 and standard deviation σ= 19.7, while using the alternative tool the results were μ=47.6 and standard deviation σ=23.2 accordingly. Due to high value of standard deviation and closeness of

mean values t test shows that there is no significant difference in the results of experiment groups.

In the third task users were asked to find the length of maximal sequence of pages, that was created during quest performing. Test group competed this task with a mean time μ=29.7 and standard deviation σ= 17.0, the results of control group were μ=29.0 and standard deviation σ=11.2. Despite that the value of mean time is lower for alternative tool, than for HistotoryLane prototype, according to t test with level of confidence p=0.93, there is no statistical difference in mean values of two groups.

The last, forth task required from users to reconstruct (recall) the browsing path that they created for answering question #5 in the quest. Test group completed this task with a mean time μ=63.1 and standard deviation σ= 43.4. Control group using the alternative tool had a mean time μ=42.6 and standard deviation σ= 28.2. Once again, due to a very high values of standard deviation the results of t test show, that with relatively high level of confidence p=0.5 there is no statistical difference in mean values of two groups.

Based on the results that are presented here it has not been possible to provide definite answers to question "Do users complete tasks related to browsing history using HistoryLane faster than using an alternative tool". However the mean time in 2 tasks out of 4 was less for HistoryLane users. Also these results did not show that HistoryLane users performed slower than users of an alternative tool, which is also a good sign.

In 3 out of 4 tasks the standard deviation value in test group was significantly higher than in control group. That fact implies, that there may be a problem with a user interface of the prototype, since the performance of users differs considerably.

Chapter 7

Conclusions and Future Work

This chapter contains the summary of the thesis work, including the results that were achieved and future work directions.

7.1 Summary of Findings and Conclusions

People intensively use Web browsers for different personal purposes and as the browsers are getting more sophisticated and provide the users with various technical features as multi-tab parallel browsing, their browsing sessions become more complicated for analysis and inspection. Browsing history inspections are inevitable part of browsing process because of the fact that users revisit most of Web pages [2].

The overall aim of this thesis was to formulate and evaluate a new approach for intuitive and effective visualization of Web browsing history that would help in presenting browsing sessions in easy for analysis way.

We started the research conducting a literature review, findings of which may be summarized through following key points:

- Modern browsers lack graphical history visualization
- Most of previous history visualization approaches ignore the parallel browsing paradigm
- Existing visualization solutions do not record the level of user activity at particular page
- Zoom-in/out feature may be very useful for browsing history visualization

Following step was to formulate main visualization principles that would be a basement for development of working prototype. Taking into account findings of literature review and visualization best practices we formulated them as follows:

- Browser windows, tabs and pages are visualized as 2D rectangles positioned from left to right in a chronological order

- To make identification of visited pages easier, entities that represent them are accompanied by favicon, thumbnail image and a page tittle
- To encode the information about category of Web page and a level of user's activity on it we manipulate colors and their brightness
- User can configure the shown parameters and the level of details

Based on these visualization principles we created a list of functional and non-functional requirements for prototype. Prototype was implemented as a Firefox extension that can be deployed in any platform. It uses standard Firefox history recording mechanism, but extends its functionality according to our requirements.

To evaluate the proposed visualization approach and answer the research questions we conducted the survey and the quantitative experiment. Here are the answers, that we have found, for research questions that were formulated in Chapter 1:

- **RQ1: Does the aggregation of browsing history into tab sections makes the perception of this history more convenient for users?** All survey participants stated that they browse multiple windows and tabs in parallel often, in addition 94% of respondents answered that they find grouping of visited pages by tabs and windows insightful. Also, providing their opinion about the prototype in open questions, users wrote that hierarchy window-tab-page is intuitive to them. Thus we can conclude that aggregation of browsing history into tabs sections makes the perception of browsing history more convenient.
- **RQ2: How to visualize collected history data ?** Absolute majority of survey respondents stated that the usage of favicons helps them to identify visited Web pages. Similar results were received about usage of titles embedded into page rectangle, that were considered useful by all but one respondent. Surprisingly survey respondents were not enthusiastic about idea of thumbnail images, 18% of them do not find this way of visualization helpful. The idea of usage of different colors for different categories of Web pages received a very heterogeneous feedback: 6% of users completely disagree with that idea, while 35% of respondent completely agree that usage of different colors helps to analyze patterns of browsing behavior. Also based on survey results we can conclude that application of vertical time lines, for making the time spent on particular Web page more obvious, is a successful idea.
- **RQ3: What are the benefits of browsing history visualization?** According to replies from the survey, visualization provides a good overview on the history, helping users to understand briefly what type of Web sites

and for how long did they visit. Another feature, that respondents of survey liked, was a highlighting of page rectangles that fit the search condition, since this feature helps to differentiate desired Web pages. Also, using the visual history, users can easily identify the sequence of visited Web pages.

Using quantitative experiment we tried to prove that HistoryLane browsing history visualization approach is more effective for specific tasks than existing alternatives. Experiment contained four tasks in which users were asked to find answers for questions about their browsing history. The results of experiment are ambiguous, since in two out of four tasks test group performed better than the control group, in other two tasks control group performed better than test group, but no statistical evidence for significant difference in mean values of two group was found in any of tasks.

7.2 Contribution

Contribution to knowledge produced by this master thesis can be divided into two main sections: new visualization ideas and results of their evaluation.

Few principles for browsing history visualization were formulated in scope of this master thesis. Among them there are:

- Representation of windows, tabs and pages using colored rectangles
- Grouping of pages by tabs, tabs by windows and their positioning as lanes
- Usage of brightness to represent the level of activity at specific Web page
- Filtering of Web pages by level of user's activity
- Option to add comments to history entries
- Combination of favicon, page title and thumbnail preview for making page entity more recognizable

Some of these principles are based on our original ideas, others use previously devised concepts, but applied differently.

By conducting the evaluation of the prototype built on base of these principles we answered three research questions, mentioned in previous section, elaborating the knowledge about users' opinion regarding browsing history visualization and its benefits.

7.3 Future Work

Since the results of quantitative experiment were ambiguous this is the most obvious direction for a future research. Another evaluation with bigger sample groups and more precise measurements would help to answer the question if HistoryLane visualization approach is better than alternative tools.

Analyzing results of quantitative experiment we payed attention, that the standard deviation of time values was high. This may be an evidence, that there are issues with the usability of HistoryLane prototype, and some users did not succeed to operate it effectively. Some survey respondents mentioned that they needed additional instructions about how to use HistoryLane. Thus design of user interface of HistoryLane should be rigorously examined, according to human interface guidelines, and simplified.

During work on HistoryLane we identified few challenges that were addressed in scope of this master thesis, but could be additionally refined in case of future work. One of these challenges is the scalability of the history visualization. In case of active browsing the width of diagram for representation of single hour of browsing was reaching 1000 pixels, to depict every opened page. Using that approach even the diagram representation for a single day becomes cluttered. In HistoryLane prototype we solved this problem using zoom in/out feature, but it should be improved. Another challenge is to make the manipulation of rectangles representing the tabs easier, so analyst could combine them in a structure, that makes the analysis of user activities more convenient.

Survey respondents also provided few ideas that should be examined in the future work. According to their inputs, users would like to have following features integrated into browsing history visualization tool:

- A graph representing the dependencies between Web pages, showing how Web pages led to each other
- An ability to moderate history entries (e.g. to delete them) as it is implemented in "regular" history mechanisms
- Graphics showing aggregated data as trends in last week/month
- More configuration options, like changing color of background or colors of Web pages categories

Appendix A

Affidavit

I confirm that I wrote this master thesis independently and on my own without using any other sources than I cited here.

Igor Chtivelband

Kaiserslautern, 11.09.2012

Appendix B

Survey Instructions and Questionnaire

Web Browsing Visual History Evaluation

To complete the survey follow this algorithm:

1. Watch on YouTube (www.youtube.com/watch?v=2J5VOtfyMRs) short tutorial about how to install the tool and how to use it.
2. Install the Firefox extension, using the .xpi file
3. Spent 15-30 minutes in a relaxed Internet browsing. Visit your favorite news portals, social networks or whatever you like to do in the Web. Don't worry, this tool doesn't report data about your activities outside and doesn't collect data about passwords/logins.
4. Use my HistoryLane tool as it was shown in tutorial to inspect your browsing history.
5. Fill up this survey!

You should evaluate the visualization ideas and not the technical realization of tool.

First some background

Web browsers are our main gateways to the Internet. We use them for activities that involve information from the Web, for example, we read articles, we learn, we listen to music, we watch videos, we share our thoughts and feelings, we write e-mails, or we chat. Considering this diversity of interactions, it is surprising how little emphasis is put on supporting insight gathering based on browsing activity and habits. As such, most Web browsers of today are still limited to a text-based browser history that has no visualization capabilities and a limited potential of filtering patterns and information. Furthermore, these histories disregard the existence of parallel navigation in multiple browser windows and tabs.

The approach that you evaluate here tries to improve the situation by depicting the parallel browsing nature of modern browsing patterns.

Question 1
Do you consider that you're browsing multiple tabs or windows in parallel often?

Completely disagree 1 2 3 4 5 6 7 8 9 10 Completely agree

Question 2
I find that grouping of visited pages into tabs and tabs into windows for visualization is useful/intuitive/insightful.

Completely disagree 1 2 3 4 5 6 7 8 9 10 Completely agree

Question 3
Representation of visited pages from left to right in chronological order is useful/intuitive/insightful.

Completely disagree 1 2 3 4 5 6 7 8 9 10 Completely agree

Question 4
Accompaniment of visited pages using favicons (those small icons from address bar) helps me a lot to identify the visited Web page.

Completely disagree 1 2 3 4 5 6 7 8 9 10 Completely agree

Question 5
Accompaniment of visited pages using thumbnails helps me a lot to identify the visited Web page.

Completely disagree 1 2 3 4 5 6 7 8 9 10 Completely agree

Question 6
Accompaniment of visited pages using page titles helps me a lot to identify the Web pages.

Completely disagree 1 2 3 4 5 6 7 8 9 10 Completely agree

Question 7
Usage of different colors for different categories of Web sites helps me to analyze patterns of my browsing behavior.

Completely disagree 1 2 3 4 5 6 7 8 9 10 Completely agree

Question 8
Vertical time lines are useful for me and help me to analyze patterns of my brows-

ing behavior.

Completely disagree 1 2 3 4 5 6 7 8 9 10 Completely agree

Question 9
List of visited pages (on the left side) is bound to drawn entities (on the right side) in an intuitive manner.

Completely disagree 1 2 3 4 5 6 7 8 9 10 Completely agree

Question 10
When I move a tab to "Tab Analysis Zone" (using double click), the representation of active_time/focus_time/visit_time using different gradation of the same color is intuitive to me.

Completely disagree 1 2 3 4 5 6 7 8 9 10 Completely agree

Question 11
An ability to expand/shrink visual entities using mouse scrolling and see additional details is useful.

Completely disagree 1 2 3 4 5 6 7 8 9 10 Completely agree

Question 12
What did you like about this visualization approach?

Question 13
What would you change or improve in this visualization approach?

Appendix C

Quantitative Experiment Quest

Here is a short challenge, please find answers for following questions using your Web browser

1. Please find in what year Napoleon 1st was born and where.
2. Please find how old he was when he invaded Russia.
3. Please find who was Tzar in Russia during that war.
4. Please find in what european capital there is a square named after that Tzar.
5. Please find in what european capital there is a bridge shown on the picture and after whom this bridge is named (Hint - it is a very famous bridge).
6. Please find what was family relation between person from previous question (the one that bridge is named after him) and russian Tzar from task #3.
7. Please find when started the last war between countries that their capitals are answers for questions #4 and #5 (one capital with a bridge and one with a square).

Figure C.1: The bridge

8. Please find when that war finished and where the surrender agreement was signed.

9. Please find what airport is a closest one to geographical point from previous question today.

Appendix D

Quantitative Experiment Tasks

Quantitative task using enhanced history prototype

This experiment evaluates how hard or easy it is to recall and reconstruct browsing patterns.

Your task is very simple:

1. Install the prototype (You should be familiar with its interface from previous experiment) using xpi located here: bit.ly/OcvEli
2. Check that it works.
3. Open document bit.ly/TNZPVg
4. Complete quest that is described there (I find it very educational) using FireFox with installed prototype as a Web browser.
5. Answer questions of this survey, using the prototype, measuring how much time in seconds did it take you to find an answer for the question. Provide this time metric, not an answer! (Example: if the question is how many days passed since your last birthday, answer how much time it took you to calculate the number, not the number).

Question 1

In what exact page did you find an answer for task #3? (Measure in seconds time to find an answer)

Question 2

How many times did you reach the answer to the task (aka the information you were looking for) right after clicking a link in a search result? (Measure in seconds time to find an answer)

Question 3
Please find the length of maximal sequence of visited pages, that you created trying to complete a single task of quest. (Measure in seconds time to find an answer)

Question 4
Please reconstruct browsing path (recall list of visited pages and write them down) that you created for completing task number 5 (the one with a bridge)? (Measure in seconds time to find an answer)

Bibliography

[1] Stuart K Card, Jock D Mackinlay, and Ben Shneiderman. *Readings in information visualization : using vision to think*. Morgan Kaufmann Publishers, San Francisco, Calif., 1999. ISBN 1558605339 9781558605336.

[2] Andy Cockburn. What do web users do? an empirical analysis of web use. *Int. J. Hum.-Comput. Stud.*, 54(6):903–922, 2001.

[3] Louis Cohen, Lawrence Manion, and Keith Morrison. *Research methods in education*. Routledge, London; New York, 2007. ISBN 0415374103 9780415374101 0415368782 9780415368780 0203029054 9780203029053.

[4] Patel Divyen. The evolution of word wide web - WWW. http://ezinearticles.com/?The-Evolution-of-Word-Wide-Web—WWW&id=869623, 2007. URL `http://ezinearticles.com/?The-Evolution-of-Word-Wide-Web---WWW&id=869623`.

[5] E. Frecon and G. Smith. WEBPATH-a three dimensional web history. In *Information Visualization, 1998. Proceedings. IEEE Symposium on*, page 310, 1998.

[6] Michael A Gallo and Bill Hancock. *Networking explained*. Digital Press, Boston, 2002. ISBN 1555582524 9781555582524.

[7] Achyut S Godbole and Atul Kahate. *Web technologies : TCP/IP, architecture, and Java programming*. Tata McGraw-Hill, New Delhi, 2008. ISBN 9780070142954 0070142955.

[8] H. Rex Hartson, Lenwood S. Heath, Robert S. Schulman, Lucille Terry Nowell, and Lucille Terry Nowell. Graphical encoding for information visualization: Using icon color, shape, and size to convey nominal and quantitative data. Technical report, 1997.

[9] Gray Hodgkinson. *Visually Representative Web History Browser*. PhD thesis, Massey University, 2007.

[10] Jeff Huang and R. W White. Parallel browsing behavior on the web. 2010. ISBN 978-1-4503-0041-4.

[11] Shaun Kaasten, Saul Greenberg, and Christopher Edwards. How people recognize previously seen web pages from titles, urls and thumbnails. In *People and Computers XVI (Proceedings of Human Computer Interaction*, pages 247–265, 2001.

[12] Gholam H. Khaksari. Direct manipulation of web browsing history. *International Journal of Computer Science and Information Technology*, 3(5): 185–192, October 2011. ISSN 09754660. doi: 10.5121/ijcsit.2011.3516. URL `http://www.airccse.org/journal/jcsit/1011csit16.pdf`.

[13] Barbara Kitchenham and Shari Lawrence Pfleeger. Principles of survey research – part 6: Data analysis, 2002.

[14] C. R Kothari. *Research methodology : methods and techniques*. Wiley Eastern : New Age International, New Delhi, 1990. ISBN 0852264771 9780852264775.

[15] Yan Li, Bo-qin Feng, and Feng Wang. Page interest estimation based on the user's browsing behavior. pages 258–261. IEEE, 2009. ISBN 978-0-7695-3634-7. doi: 10.1109/ICIC.2009.72. URL `http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=5169589`.

[16] Richard G Lomax. *An introduction to statistical concepts*. Lawrence Erlbaum Associates Publishers, Mahwah, N.J., 2007. ISBN 9780805857399 0805857397 9781410616425 1410616428.

[17] Richard Lowry. *Concepts and Applications of Inferential Statistics*. Vassar, Poughkeepsie, NY, 2007.

[18] M. Mascoet. Interaction and visualization supporting web browsing patterns. In *Information Visualisation, 2001. Proceedings. Fifth International Conference on*, page 413418, 2001.

[19] Marissa Mayer. One fish, two fish, red fish, blue fish | official google blog. http://googleblog.blogspot.de/2008/06/one-fish-two-fish-red-fish-blue-fish.html, 2008. URL `http://googleblog.blogspot.de/2008/06/one-fish-two-fish-red-fish-blue-fish.html`.

[20] Matthias Mayer. *Visualizing Web Sessions: Improving Web Browser History by a Better Understanding of Web Page Revisitation and a New Session- and Task-Based, Visual Web History Approach*. PhD thesis, University of Hamburg, 2007.

[21] N. Milic-Frayling, R. Sommerer, and K. Rodden. WebScout: support for revisitation of web pages within a navigation session. In *Web Intelligence, 2003. WI 2003. Proceedings. IEEE/WIC International Conference on*, page 689693, 2003.

[22] T. Nadeem. A study of three browser history mechanisms for web navigation. In *Information Visualisation, 2001. Proceedings. Fifth International Conference on*, page 1321, 2001.

[23] Jacob Nielsen. Reviving advanced hypertext (Jakob nielsen's alertbox). http://www.useit.com/alertbox/20050103.html, March 2005. URL `http://www.useit.com/alertbox/20050103.html`.

[24] RIA Novosti. Number of internet users worldwide tops 2 billion. http://en.rian.ru/world/20120120/170857236.html, 2012. URL `http://en.rian.ru/world/20120120/170857236.html`.

[25] S. P Reiss and G. Eddon. Visualizing what people are doing on the web. In *Visual Languages and Human-Centric Computing, 2005 IEEE Symposium on*, page 305307, 2005.

[26] Jakub Simko, Michal Tvarozek, and Maria Bielikova. Semantic history map: Graphs aiding web revisitation support. pages 206–210. IEEE, August 2010. ISBN 978-1-4244-8049-4. doi: 10.1109/DEXA.2010.55. URL `http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=5592075`.

[27] Linda Tauscher and Saul Greenberg. Design guidelines for effective www history mechanisms. In *In Proc. of Workshop on Designing for the Web: Empirical Studies Microsoft Corporation*, 1996.

[28] M. Van Kleek, B. Moore, C. Xu, and D. R Karger. Eyebrowse: real-time web activity sharing and visualization. In *Proceedings of the 28th of the international conference extended abstracts on Human factors in computing systems*, page 36433648, 2010.

[29] Elina Vartiainen, Virpi Roto, and Janne Kaasalainen. Graphical history list with multi-window support on a mobile web browser. pages 121–129. IEEE, June 2008. doi: 10.1109/ICIW.2008.13. URL `http://ieeexplore.ieee.org/lpdocs/epic03/wrapper.htm?arnumber=4545602`.

[30] w3school. Browser statistics. http://www.w3schools.com/browsers/browsers_stats.asp, 2012. URL `http://www.w3schools.com/browsers/browsers_stats.asp`.

[31] Colin Ware. *Information visualization : perception for design*. Morgan Kaufman, San Francisco, 2000. ISBN 1558605118 9781558605114.

[32] A. Woodruff, A. Faulring, R. Rosenholtz, J. Morrsion, and P. Pirolli. Using thumbnails to search the web. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, page 198205, 2001.

[33] Chien-Fu Wu and Michael Hamada. *Experiments : planning, analysis and parameter design optimization*. Wiley, New York; Chichester, 2009. ISBN 9780471699460 0471699462.

[34] T. Yamaguchi, H. Hattori, T. Ito, and T. Shintani. On a web browsing support system with 3d visualization. In *Proceedings of the 13th international World Wide Web conference on Alternate track papers & posters*, page 316317, 2004.