Ludwig Maximilian University of Munich
Institute of Computer Science
Media Informatics and Human-Computer Interaction Groups
Prof. Dr. Heinrich Hußmann

Master's Thesis

Utilizing Users' Web Browsing and Search Behavior to Improve Website Revisitation

Sven Unnewehr
s.unnewehr@campus.lmu.de

Date:	October 27, 2015
Duration:	May 1, 2015 to October 31, 2015
Company:	CLIQZ GmbH
Supervisor at Company:	Dr. Sean Gustafson
Supervisor at University:	Prof. Dr. Florian Alt

Abstract

Internet users regularly need to re-find information or content that they looked at in the past. In some cases, these revisitations take place weeks after the initial visit. Long-term revisitations, also called rediscoveries, are often time-consuming, prone to failure and require high mental effort. Existing research showed that current browsers poorly support this activity requiring users to rely on less efficient strategies, such as re-creating queries or re-tracing previous browsing paths, to find the desired information.

In two formative studies, I confirmed the existing findings and showed that, on average, rediscoveries take about the same time as the initial search for the information, users often fail because of trouble identifying pages and users are unable to make use of contextual memories.

These insights led me to the development of the *CLIQZ Browsing History*, which acts as a replacement for the browser's history list. Common user behaviors and memories are directly supported by grouping the history into sessions, by showing context and by providing a searchable query history. Additionally, users are able to explore previous browsing paths and recognize pages using mouseover previews.

To evaluate the developed tool, I conducted an evaluation, which confirmed the benefits of the underlying concepts with a promising performance increase after continued usage and users needing significantly fewer page visits for successful rediscoveries.

Zusammenfassung

Internetnutzer stehen regelmäßig vor der Herausforderung, Internetseiten wiederzufinden, die sie in der Vergangenheit besucht haben. Diese Wiederbesuche finden oft mehrere Wochen nach dem ursprünglichen Besuch statt. Langfristige Wiederbesuche sind häufig zeitaufwendig, mühsam und fehleranfällig. Frühere Studien haben gezeigt, dass solche Wiederbesuche kaum von Browsern unterstützt werden und Nutzer stattdessen die Informationen erneut suchen oder ihren Browsing-Pfad wiederholen.

Ich habe diese Feststellungen in zwei Studien bestätigt; im Durchschnitt dauert das Wiederfinden einer Internetseite genauso lange wie die ursprüngliche Suche. Nutzer haben Schwierigkeiten, Internetseiten zu identifizieren, können keine kontextbezogenen Erinnerungen nutzen und sind oft nicht in der Lage, die gesuchte Seite zu finden.

Diese Erkenntnisse haben mich dazu bewegt, die *CLIQZ Browsing History* als Alternative zur browsereigenen Gesamtliste des Internetverlaufs zu entwickeln. Gebräuchliches Nutzerverhalten und Erinnerungen werden direkt unterstützt, indem der Internetverlauf in Sitzungen gruppiert, Kontext angezeigt und ein durchsuchbarer Suchverlauf bereitgestellt wird. Zusätzlich können Nutzer frühere Browsing-Pfade nachvollziehen und Internetseiten mithilfe einer Mouseover-Vorschau einfacher identifizieren.

Um die Effektivität der *CLIQZ Browsing History* zu bewerten, habe ich eine Evaluation durchgeführt, die die zugrunde liegenden Konzepte befürwortet. Ich habe eine Performance-Steigerung nach längerer Nutzung beobachtet und die Anzahl der Seitenaufrufe für erfolgreiche Wiederbesuche wurde signifikant reduziert.

Scope

In recent years, Internet usage has continued to grow and more websites are visited than ever before [19]. People use the Internet for many different purposes and on many different devices. When looking at information and content on the Internet, there is also a need for re-finding this information. This process is often called rediscovery, i.e., the user is actively looking for a web page that he has visited before. Rediscoveries do not include regularly visited websites, but rather information or content that was only seen once and requires some effort to find again. These rediscoveries make up about 7% of all page visits on average and are often time-consuming, prone to failure and require high effort [37].

As rediscoveries are poorly supported by browsers, many potential solutions have been explored, e.g. YouPivot [18] or SearchBar [36]. All of these approaches focus on specific areas of revisitation, e.g. re-searching content using a search engine or combining the browsing history with contextual information. However, no tools exist of today that support every aspect of rediscovery and offer a simple and easily usable interface.

The aim of this project is to create a history interface that makes it possible to rediscover content using a range of strategies. As the result is potentially integrated into the CLIQZ browser extension, this work will also put a strong focus on usability and interaction design. The final approach is not supposed to perform better for special use cases where solutions have already been explored. Instead, the tool should be easily usable and make it simple to rediscover web pages by supporting user strategies, behaviors and memories.

Tasks

- Comprehensive analysis of related work
- Confirm previous findings by looking at today's user behavior
- Design and development of a prototype for a history interface
- Evaluation of the developed prototype

Ich erkläre hiermit, dass ich die vorliegende Arbeit selbstständig angefertigt, alle Zitate als solche kenntlich gemacht sowie alle benutzten Quellen und Hilfsmittel angegeben habe.

München, December 4, 2015

.....

Acknowledgments

First of all, I would like to express my deepest gratitude to my supervisor Dr. Sean Gustafson for his support, guidance, feedback and suggestions throughout this thesis. His great knowledge, experience and insightful comments always helped me during the development of my ideas and writing of the thesis. Furthermore, I would like to thank Prof. Dr. Florian Alt for making this cooperation possible and always giving valuable suggestions and feedback.

I want to thank the four participants of my initial user study: Richard, Valerie, Humera and Harry. Your thoughts and problems sparked many of the ideas that guided me throughout this thesis.

Thanks to Christian for testing my questionnaire and giving me great insights.

My sincere thanks to my colleagues who tested my prototype and my evaluation: Stephie, Tobias, Felix and Richard. Your invaluable feedback helped me immensely.

Thanks to Lucian for pushing my prototype and evaluation to a great number of CLIQZ users.

Thank you to all participants of my evaluation for providing valuable insights and feedback. Special thanks to all external CLIQZ users who sacrificed their free time to help a stranger over the Internet.

Thanks to CLIQZ and all my other colleagues for making this thesis possible, providing financial support and also offering a price for participants of my evaluation. I enjoyed my time at CLIQZ and am very grateful for this great opportunity and everything I learned.

Finally, I would like to thank my parents for their moral and financial support during my whole studies.

Contents

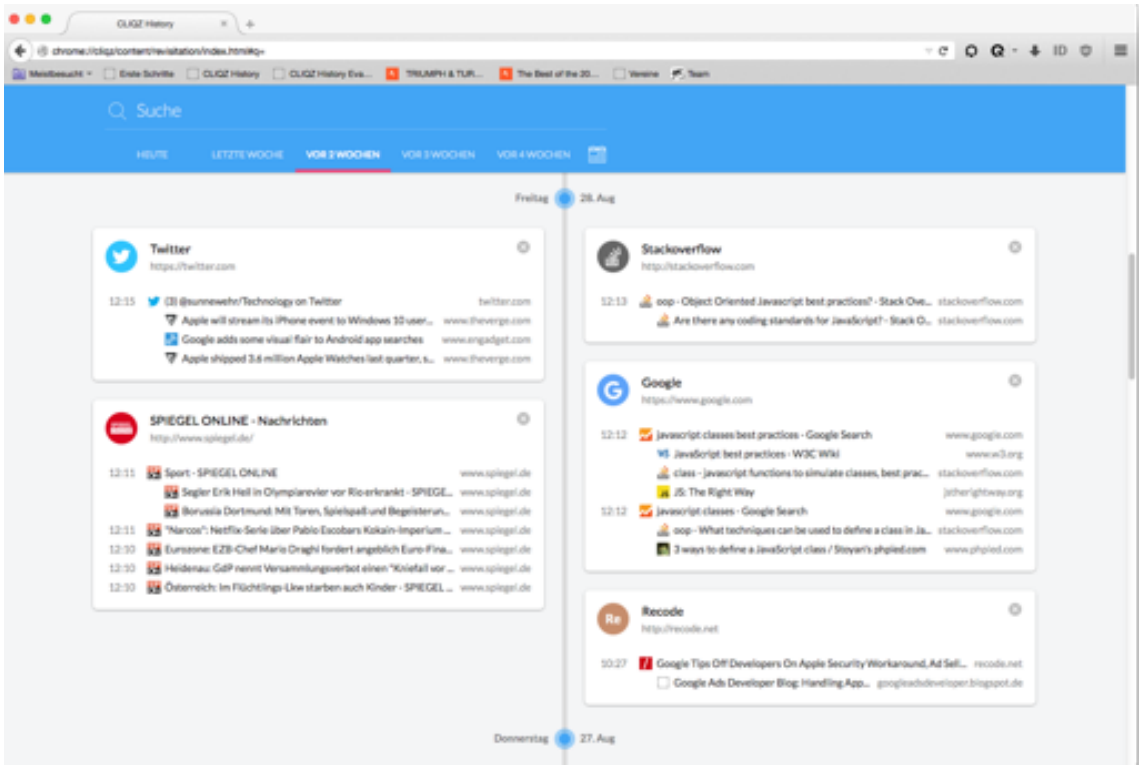

Figure 1.1: The CLIQZ Browsing History shows the user's history and allows to search, filter and quickly navigate visited websites. The interface supports users' behavior and memory by providing context, previews and search suggestions.

1 Introduction

This thesis examines how Internet users revisit websites and web pages, what strategies they use and what problems they encounter. To address these problems and to support website revisitation, I built *CLIQZ Browsing History*, a history search tool that shows an overview of the user's browsing history and provides several features which support the identified strategies and behaviors (see Figure 1.1).

In recent years, Internet usage has continued to grow and more websites are visited than ever before [19]. People use the Internet for many different purposes and on many different devices. When looking at information and content on the Internet, there is also a need for re-finding this information. This process is often called rediscovery, i.e., the user is actively looking for content or information that he has visited before. Rediscovery does not include regularly visited websites, but rather pages that were only visited once and require some effort to find again. These rediscoveries make up about 7% of all page visits on average [37]. This thesis takes a detailed look at rediscovery, including the problems and strategies involved and how to make this process easier, faster and more successful.

This work resulted from a cooperation with CLIQZ, a company located in Munich that set itself the goal to make it easier for people to navigate the Internet. Their main product is a Firefox extension that shows relevant information while the user is typing in the URL bar, e.g., search results, weather or news articles (see Figure 1.2). Users save time by not having to go to separate websites where they would have to search for the information.

Furthermore, the CLIQZ browser extension shows websites from the user's browsing

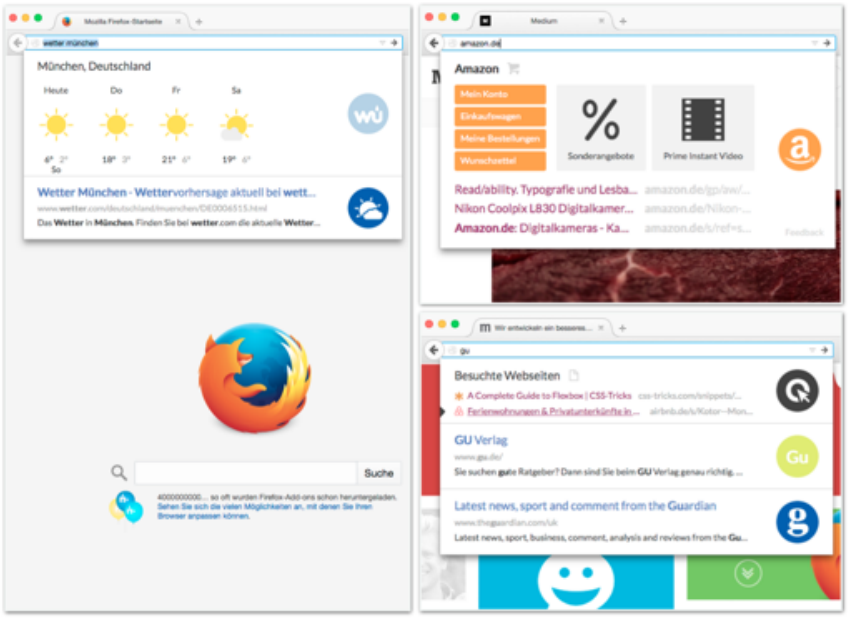

Figure 1.2: The CLIQZ browser extension shows relevant information in a dropdown below the address bar, e.g., weather, website links or browsing history.

history and makes it possible to use the address bar to search for specific web pages that were visited before. Most users use the browsing history mostly for autocompletion and quick access to regularly visited websites. Rediscovering web pages without knowing the URL or title of the website is difficult, because the address bar acts as a filter for the browsing history. If the user does not remember anything that he can use as a query, he has to fall back to other search strategies. This problem is not exclusive to the CLIQZ extension; all browsers currently use very similar mechanisms to search the browsing history.

These observations are confirmed by Obendorf et al., who concluded that "Neither browser history nor bookmarks seem to be reliable tools for long-term rediscovery. Instead, users re-searched and re-traced the Web for the desired information." [37]. Although rediscovery makes up around 7% of all visits, there is no browser support and many people struggle at re-finding web pages. Rediscovery offers big potential for improvement which motivated me to take a deeper look at this topic.

1.1 Benefits and Contributions

The major contributions of this thesis are:

1. A quantitative and qualitative analysis of how users rediscover web pages and websites from their browsing history, what strategies they use and what problems they encounter.

2. The design and development of *CLIQZ Browsing History*, a history interface which provides an overview of the user's history and supports the observed behavior and strategies.

3. A comprehensive evaluation that compares the developed history tool to traditional rediscovery strategies.

1.1.1 Understanding Revisitation Behavior

The user analysis consists of two parts: The first study looks at quantitative data from 26 users of the CLIQZ extension. This data shows that revisitation takes, on average, about

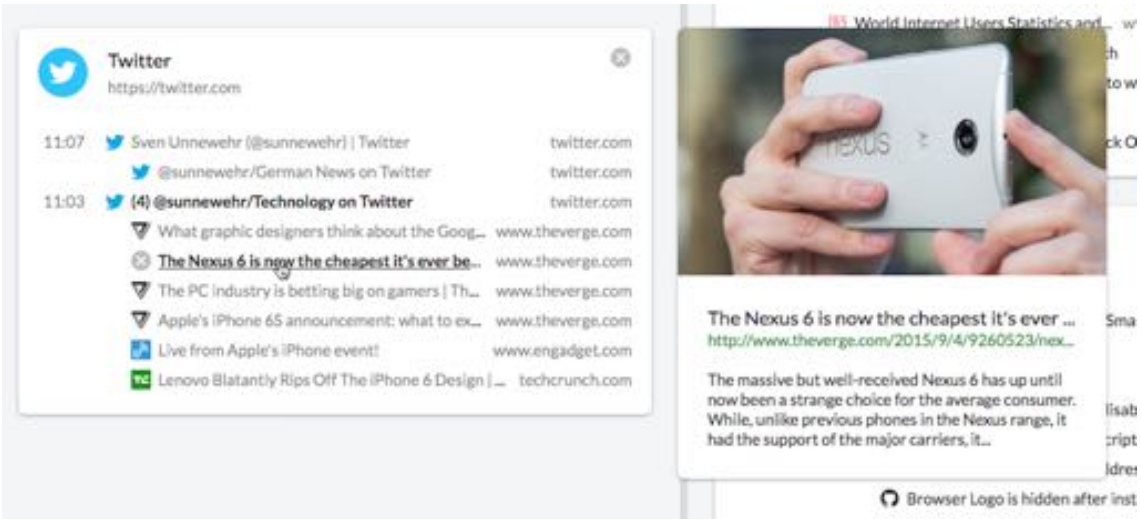

Figure 1.3: CLIQZ Browsing History shows context, i.e., pages that were visited in sequence and also provides previews and descriptions to easier identify the sought web page.

Figure 1.4: CLIQZ Browsing History offers a search feature with suggestions. This can be used to filter searches the user has done in the past and also search for specific web pages.

as long as the original visit. *Having visited a web page provides no time saving when going back to it.* In other words, rediscovering pages represents a new effort that is likely to take the same amount of time. There are no browser tools that support common memories and users have to rely on inefficient and often unsuccessful strategies. Secondly, four interviews were conducted, which focused on specific problems that users encounter but also which strategies they use when rediscovering web pages. The participants encountered a range of problems and often failed to re-find web pages from their browsing history. Mostly, they did not remember enough information to start a specific search or could not make use of their memories. Usually participants re-created the path to the website, i.e., if they were looking for a search result, they tried to repeat the same search to get to the result.

1.1.2 CLIQZ Browsing History Prototype

The prototype is based on the findings from the user analysis. I identified several problems that users encounter while rediscovering websites and web pages:

1. Browser tools that are designed for revisitation are mostly unknown or unused and have usability problems.

2. Users are unable to use contextual information to initiate a query as browser tools require the user to remember the title or URL.

3. Common strategies, e.g., re-searching or keyword searching, are time-consuming, error-prone and poorly supported by browsers.

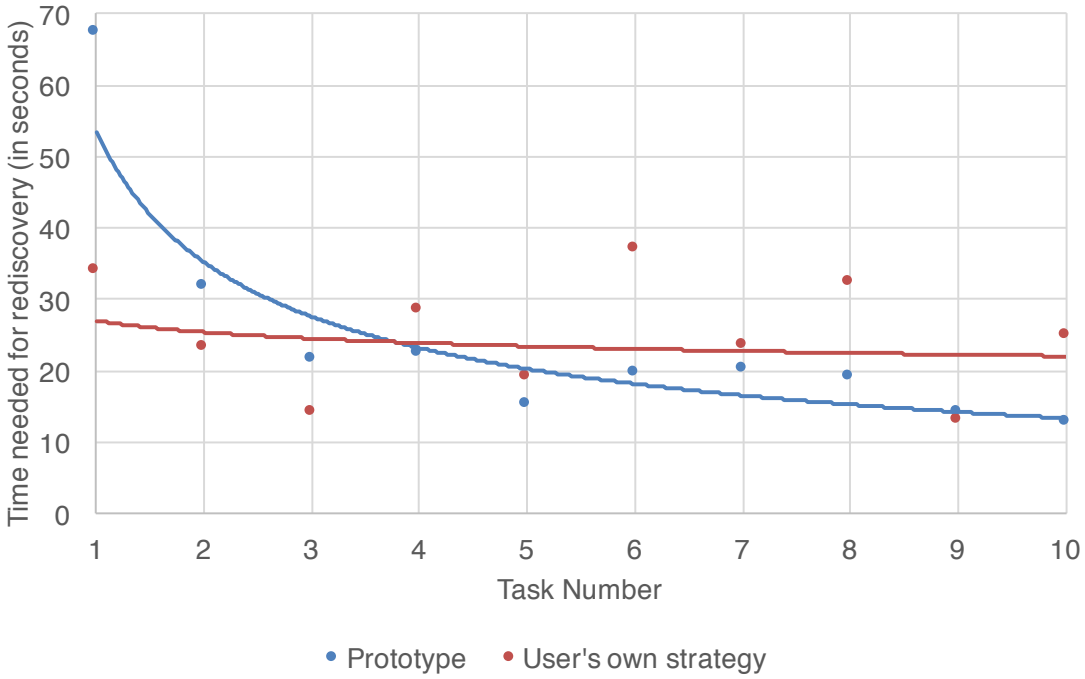

Figure 1.5: Users who used the CLIQZ Browsing History for rediscovery performed poor in the first two tasks and performed better for later tasks. This learning effect cannot be observed for the second condition as users used familiar strategies.

4. Identifying the correct web page is often difficult and requires high effort, especially when many similar pages were visited, e.g., when looking at search results or articles on a shopping website.

I developed *CLIQZ Browsing History*, a history interface that shows an overview of the user's browsing history and offers several features that consider these issues. Users are able to see context, filter their search queries (see Figure 1.4), jump to specific time frames and see previews (see Figure 1.3) to easily identify web pages.

1.1.3 Evaluation of the CLIQZ Browsing History

To evaluate the developed prototype, 28 users of the CLIQZ extension took part in a study that compared the users' preferred rediscovery methods to the prototype. The results show that users are slightly faster and more successful when rediscovering web pages using the CLIQZ Browsing History. These differences were not significant, however I observed a promising performance increase in later trials (see Figure 1.5). A future study of long-term users could show further benefits. The average number of page visits needed to rediscover pages was reduced significantly from 2.0 to 1.32. By analyzing the interaction with the prototype, I found that users are still able to utilize familiar strategies while requiring less mental effort. All in all, the evaluation results support the underlying concepts.

Finally, the usability of the CLIQZ Browsing History was determined using the System Usability Scale (SUS), which resulted in a score of 75 out of 100, which is considered above average [48].

1.2 Structure of Thesis

The remainder of the thesis is structured as follows: Chapter 2 gives an overview of related work in the fields of browsing behavior, browsing history, website recognition and visualization, as well as psychological backgrounds. Chapter 3 deals with the quantitative and qualitative user analysis and its implications. The prototype, its interaction, interface and features are described in Chapter 4. Next, Chapter 5 explains the evaluation design, execution and analysis. The last chapter wraps up the thesis and lists recommendations and learnings as well as the potential for future work.

2 Related Work

The starting point for this thesis is the rediscovery of web pages and websites. There are several connected fields, such as general browsing behavior, website recognition, browser usage and also psychological aspects from other areas, e.g., way-finding and navigation. Starting with a general overview of how people browse the Internet, this chapter will highlight the connections between these fields and how they affect the rediscovery and revisitation of websites.

2.1 Browsing Behavior

When browsing the Internet, users mostly go to websites that they have visited before. Cockburn et al. found that the revisitation rate, i.e., the proportion of websites that were previously visited, is about 80% [12]. On the other hand, only one in five websites are visited for the first time.

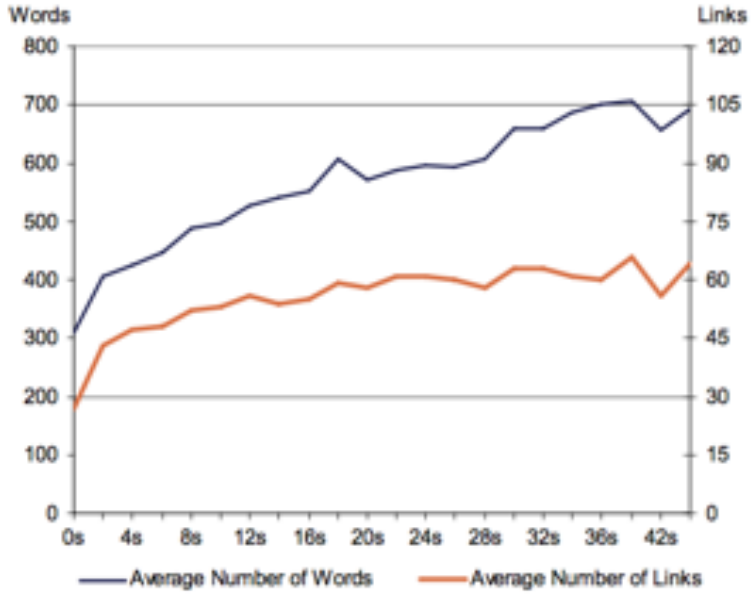

Figure 2.1: Relationship between stay time and page contents [50].

Furthermore, users do not stay on a specific web page for a very long time, 50% of websites are shown for less than 12 seconds [51]. The short stay time combined with a high number of visited websites makes remembering and especially re-finding visited pages difficult. These findings are confirmed by Weinreich et al., who determined a median display time of about 10 seconds. This study observed a lower revisitation rate of about 45% [50]. The assumption that the low stay time is caused by shortly revisited pages turned out to be false. Therefore, the authors concluded that this behavior is caused by users generally scanning websites, i.e., briefly looking at the structure of the website and not explicitly reading specific content. The number of words and links on a page are related to the stay time (see Figure 2.1). Thus it is concluded that the scannability and structure of a web page is important for usability and what users remember about pages they visit [50].

In recent years, the usage of the browser's back button has declined. This is mostly attributed to a rising number of interactive pages [50]. Users also tend to use multiple windows or tabs and follow several browsing paths in parallel. These factors combined

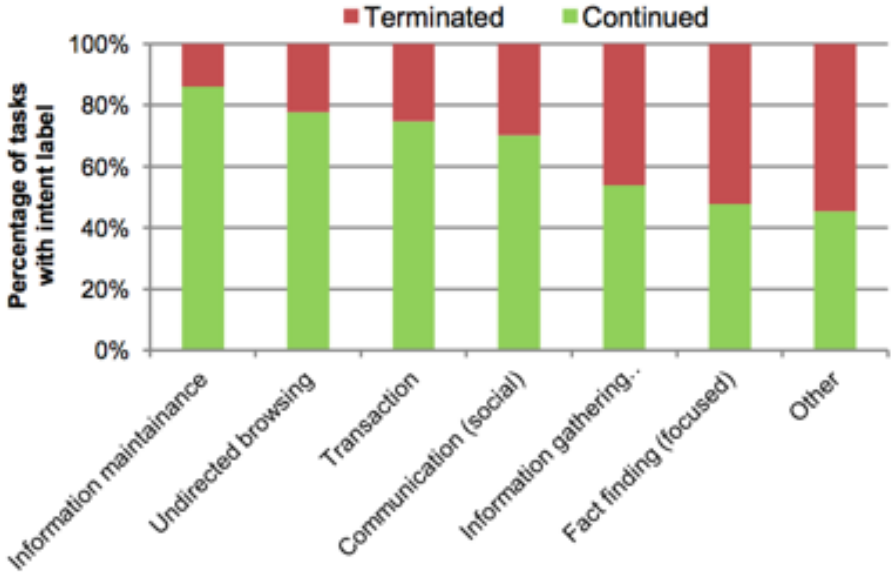

Figure 2.2: Users are very likely to continue previous browsing tasks [3].

make it very difficult for users to backtrack: Interactive web applications mostly do not support the back button and multiple open pages all have separate history stacks. The authors conclude that current browsers still use a design that was not intended for today's browsing behavior [50].

People use the Internet for many different purposes, such as looking for information, research, browsing, transactions or communications. All these tasks involve different usage patterns [30]. While some tasks are mostly done within a single browsing sessions, there are also more complex tasks that last over multiple sessions, such as information gathering. In a study by Morris et al., 83% of users reported that they have done multi-session browsing in the past [36]. In these cases, people often need to return to previously visited pages and often use different tabs and look at a great number of pages [3, 30]. Furthermore, users are very likely to go back to previous tasks (see Figure 2.2), which increases the need for re-finding pages [3]. These use cases all require different browser tools, such as the back button for single sessions or the browsing history for multi-session tasks [30]. However, these tools do not provide sufficient support and were not designed for these tasks [11, 32]. In addition to multi-session tasks, users also do different tasks in parallel, i.e., multitasking [49]. On average, users have 2 – 8 ongoing tasks simultaneously [49]. One important aspect of session based browsing is to identify these sessions. There are different approaches that try to automatically detect sessions [55], for example by grouping search queries within a certain time frame.

Searching is an important part of browsing, especially for more complex tasks. An average search session consists of 2.15 queries, with 3.16 words per query [39]. Users often modify their queries by adding, deleting or changing search terms. Furthermore, users look at many different results and often go back and forth between them [10, 24]. Returning to a specific search or result is made difficult by the number of different queries, viewed results and also other tasks done in parallel [39]. To continue tasks or re-find information users often repeat searches, which is made difficult by these behaviors.

All in all, today's browsing behavior is characterized by a great number of visited pages. Users scan pages, rather than reading them in detail which results in average page show times of a few seconds. Moreover, people do complex tasks on the web which often consist of several browsing sessions. Searching often involves a sequence of queries and

	Catledge & Pitkow [6]	Tauscher & Greenberg [32,33]	McKenzie & Cockburn [8,24]	This Study [36,37]
Date of Study	1994	1995-1996	1999-2000	2004-2005
Method	Instrumented XMosaic	Instrumented XMosaic	Daily Netscape History & Bookmark File Backups	Web Proxy & Instrumented Firefox
Data Captured	Visits & User Actions (34 Types)	Visits & User Actions (32 Types)	Visits & Bookmark History	Page Stats, Visits & User Actions (76 Types)
Length (days)	21	35-42	119	52-195 (a=105, fi=109)
# Users	107	23	17	25
Link Events	45.7%	43.4%	–	43.5%
Back Events	35.7%	31.7%	–	14.3%
Form Submit	–	4.4%	–	15.3%
∑ Direct Access	12.6%	13.2%	–	9.4%
Recurrence Rate	61% (reported in [35])	58%	81%	45.6% (43.7%, see text)
Type of Users	100% CS	100% CS	100% CS students	64% CS, 36% other academics
Visits	31,134	19,000	84,841	137,272
URIs			17,242	65,643
Visits / User		>300	281 – 23,973	912 – 30,756
Ø Visits / Day	14	21	41	89.8 (per active day)

Figure 2.3: Comparison of different studies on revisitation. The revisitation rate ranges from 45 to 81%, while most websites are visited by back events or direct access [37].

looking at different results. Looking at such a high number of pages in a short time also makes it necessary to revisit pages, e.g., to continue a previous session or to re-find specific information. These rediscoveries make up to 11% of all visits and current browsers offer little support for these revisits [37, 50]. The next chapter will explain how people currently deal with this problem, what solutions were explored and how these findings can provide better browser support in the future.

2.2 Revisitation

Revisiting websites and web pages is an important part of browsing. Depending on the study, the revisitation rate ranges from 45% [50] to 80% [12]. Most of these revisits are done by using the back button or directly accessing a website (see Figure 2.3). These direct accesses are usually websites that are visited regularly, e.g., mail, social networks or banking.

The revisitation rate also depends on the type of website. Websites that have “hubs”, e.g., categories on news websites, often have a higher revisitation rate than search-based websites like Google or dictionaries [37]. Also, most users have a small pool of websites that make up the majority of page visits [42]. In a study by Obendorf et al., the proportion of visits of users’ top ten websites ranges from 38% to 90% [37].

Another important aspect of revisitation is the time between the first visit and the revisit. This revisitation pattern can be used to characterize a website (see Figure 2.4). Fast revisits, i.e., under an hour, represent shopping or hub sites, while slower revisits mostly happen on infrequently visited sites like entertainment or free-time activities [2, 37, 43].

Adar et al. found that fast revisits are likely to be preceded by a visit from the same domain [2]. Therefore, these revisits are mostly done by using the back button or navigational links on the website. Medium-term revisits that occur between an hour and up to a day, mostly consist of mail, bank, news or sport pages. These websites are visited very frequently and are usually accessed directly, e.g., by entering the URL in the address bar. This category of visits includes the users’ most visited and popular websites and

Cluster Group	Name	Shape	Description
Fast Revisits (< hour) 23611 pages	F1		Pornography & Spam, Hub & Spoke, Shopping & Reference Web sites, Auto refresh, Fast monitoring
	F2		
	F3		
	F4		
	F5		
Medium (hour to day) 9421 pages	M1		Popular homepages, Communication, .edu domain, Browser homepages
	M2		
Slow Revisits (> day) 18422 pages	S1		Entry pages, Weekend activity, Search engines used for revisitation, Child-oriented content, Software updates
	S2		
	S3		
	S4		
Hybrid 3334 pages	H1		Popular but infrequently used, Entertainment & Hobbies, Combined Fast & Slow

Figure 2.4: Websites can be characterized by their revisitation patterns. Shopping websites have higher revisitation rates than hobby or activity sites [2].

therefore makes up the majority of all visits [2]. Lastly, there is also a group of pages with slow revisitation patterns with intervals longer than a day. This category differs from the faster groups because websites are often accessed by search engines instead of direct access. Pages in this group come from various categories, such as hobbies, movies or infrequent information gathering [2].

Slow revisits with a revisit time of more than a day also include so-called rediscoveries. These are revisits of websites that were only visited once or a few times at most. The following section will focus on the process of rediscovering websites, including what people remember about websites, what strategies they use for rediscovery and what problems they encounter.

2.3 Rediscovery

Rediscovery is a type of long-term revisit where the time since the first visit is longer than one day, often longer than a week. Therefore, it often requires some effort to find these pages again. The rediscovery rate varies from user to user, but is mostly in the range of 1 to 11% [37]. Despite this relatively low rate, finding these pages is very important for users [37]. In comparison to shorter revisits, the number of visits by direct access is much lower while the number of visits by link is very high (see Figure 2.5). Instead of entering the URL directly, users fall back to various other strategies, such as re-searching or re-tracing [37]. As there is little browser support for rediscoveries, this process often involves a series of problems for users. [4].

2.3.1 Strategies & Problems

When users rediscover web pages, they use a variety of strategies. Short to medium-term revisits are usually done by direct access. In these cases, users know the URL because they frequently visit these websites and have little trouble going there again. However, long-term revisits are often done for pages that were only visited once. Therefore, users have trouble remembering the domain and especially the URL because it might be a deep link on a rarely visited website. Depending on what the users remember about the initial visit, there are different strategies that can be utilized [4].

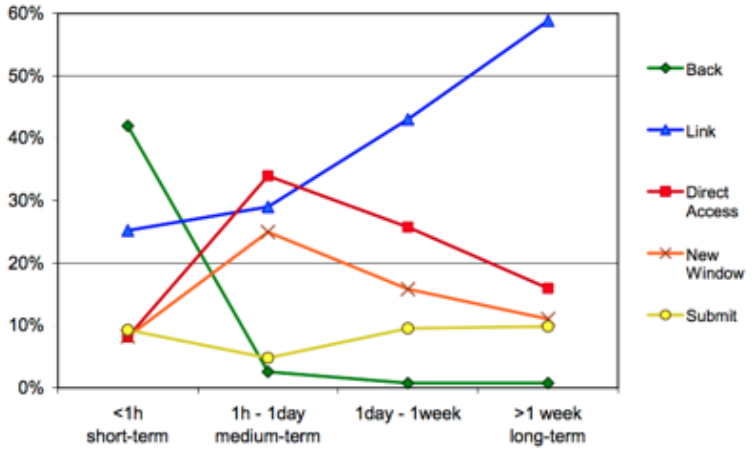

Figure 2.5: Slower revisits are rarely done by direct access. Users fall back to other strategies, such as re-searching and re-tracing the desired page [37].

In a study by Aula et al., 14 different strategies were analyzed (see Figure 2.6). Out of these strategies, only three were features provided by the browser, from which two (bookmarks and history tool) were designed for saving and revisiting websites. However, neither bookmarks nor the history tool are frequently used and both have usability problems, especially for rediscovery [9,25,31,37]. The browser's history tool is a mostly unknown and very rarely used feature. The usage rate ranges from 0% [9] to around 1% [4]. Furthermore, it requires the user to enter a specific query which filters the history by comparing it to the title and URL of visited websites. As users often do not remember the title, URL or even domain, the history tool is rarely useful for rediscoveries [4]. Alternatively, it is possible to manually look through the history list to find a specific website. This is a very time-consuming and rarely successful strategy because most users have long, cluttered history lists which make it difficult to identify the correct pages [4]. Current browsers also allow to search the history by using the address bar. This feature provides a quick access to the search function in the history tool and therefore has similar problems, i.e., the user has to remember the URL or title of the desired page.

Bookmarking is another browser feature that potentially helps users to rediscover websites by saving and organizing visited pages. However, bookmarks are only used by around 3% [25] of users, mostly because bookmark management is difficult and users need to save the page the moment they visit it. When a page is visited, users often do not know whether it is worth revisiting later. Making this decision requires mental effort and having too many bookmarks involves other problems. Users with less than 35 bookmarks tend to use no folders [1]. Above this threshold, i.e., when an unorganized list becomes unmanageable, people start to create and manage a folder hierarchy. At this point, managing the hierarchy, naming folders and retrieving bookmarks becomes more and more difficult which also decreases the likelihood for users to add new bookmarks. All in all, bookmarks require decision making and maintenance, they become cluttered and make finding pages difficult [22]. There are approaches that make saving bookmarks easier, for example by automatically sorting them into the right folders [41] or visualizing them [34], but these are currently not used in any browsers [30].

Research shows that the browser tools which were designed for rediscovery are inefficient and have usability problems (see above). This raises the question which rediscovery strategies are used successfully by users. Bruce et al. found that the strategy used by 66% of users is to "do nothing to save but search again to re-access" [9]. These findings are confirmed by other studies where the most common strategies were keeping the page open in a tab or window or searching again [4,25,37,39], while bookmarks and the history were rarely used. Furthermore, common strategies also included saving the URL in an email,

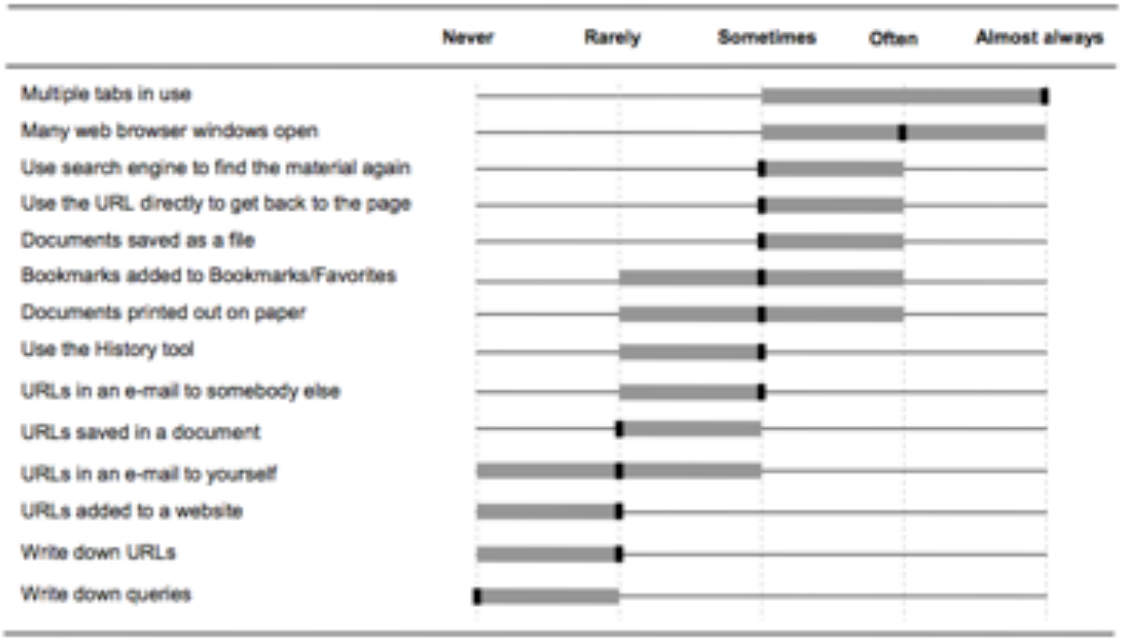

Figure 2.6: Strategies that are used for re-accessing information on the Internet [4].

Search Stage / Sessions	Finding	Re-Finding
Total number of search sessions	59	57
Total number of search queries	122	144
Total number of search terms	199	239
Average length of search terms (Chinese characters)	2.78	3.21
Average length of search terms (English words)	1.94	2.16
Average number of queries per session	2.15	2.67
Average number of search terms per query	1.63	1.83

Figure 2.7: When re-finding information on the Internet, users try to re-create their initial searches. This approach requires a higher number of queries and modifications to search terms [39].

writing it down, sending it to someone else or even printing the page (see Figure 2.5) [21].

In contrast to short and medium-term revisits, rediscoveries involve a range of strategies that might be even used in conjunction. Pu et al. analyzed in which cases users use specific strategies and how these perform [39]. Users were presented with the task to find a specific page or piece of information and asked to re-find it a week later, i.e., to rediscover it. The initial stage was mostly straight forward: Participants used either a search engine or the search function on websites to search for keywords. However, the rediscovery stage depended on the participants' memory and involved different strategies. When users only remembered information about the target and not how they got there, they initiated a new search using this information. On the other hand, users tended to use a different strategy when they remembered waypoints or what they searched for. In these cases they utilized a orienteering strategy where they tried to re-create the exact query or browsing path that they used in the initial stage [37, 39]. Re-creating a query is a difficult process because users generally do not remember the exact search terms. Therefore, they need several search attempts and modifications to their query to get the same search results (see Figure 2.7). As a result, re-finding information and web pages requires a higher number of queries and search terms. This problem is amplified by the fact that searching is often an iterative process where many different queries are used and users look at various results [4, 38]. The time needed to rediscover content is about the same as initially finding it [37, 39], but there is also a failure rate of about 7% [9].

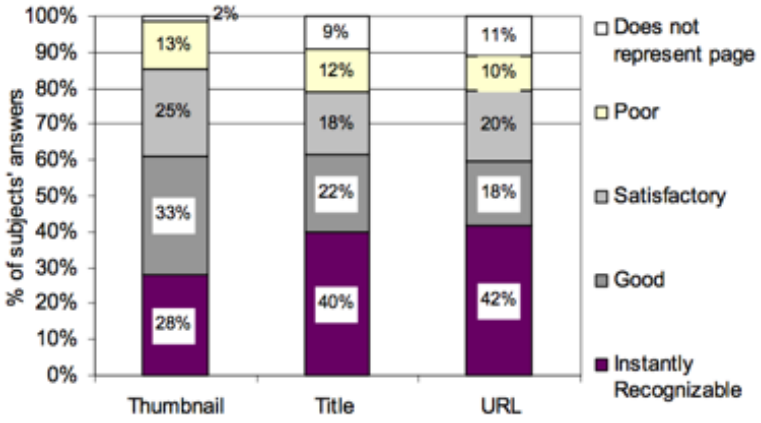

Figure 2.8: When used in isolation, thumbnails perform slightly better than title and URLs for website recognition [40].

Wen et al. identified the same strategies in a similar study where users also had to retrieve previously visited pages [52]. In this study, all participants used some kind of re-tracing: They started at a page they remember from their browsing session and tried to re-create their initial browsing path. Users like to return to a known starting point [6]. Therefore, they first went to Google when they wanted to re-find search results. Participants stated that they would rather use a longer and predictable browsing path rather than searching for the target directly. Re-tracing web content was generally the preferred method, but only had a success rate of about 20% [6]. Obendorf et al. identified several problems when re-finding pages using a search engine: Users reported that they had trouble remembering the query and rapidly changing result pages made it difficult to identify the correct pages [37]. Search engines do not provide a searchable list of previous queries which made reproducing the exact query difficult. Even if users were able to re-create the exact query, the search results could have been different and therefore difficult to recognize [44]. Another problem is the lack of browser support, i.e., the only feature that helps users re-tracing is the highlighting of visited links. This feature does not help users to find a starting point and also gets less useful for bigger browsing sessions. When users visit many links on a given page, they will still have trouble identifying the correct page and re-creating the path [37].

2.3.2 Memory and Website Recognition

To provide better support for rediscovering websites, it is important to identify what users remember about websites and browsing sessions. The average show time of a website is less than 12 seconds [51] and users tend to scan pages rather than read specific text [50]. Therefore, people generally have a very good visual memory of the structure of a web page, while they often do not remember the exact title or URL, but rather keywords from the page [7]. Further attributes that are often recalled besides visual elements are time of visit and also associated events [7]. To support the users' visual memory, thumbnails have been explored as an alternative to text representations [33]. Kaasten et al. have discovered that thumbnails perform slightly better than title and URL to recognize visited pages (see Figure 2.8). Further research shows that thumbnails work especially well when combined with other information such as title or URL [5, 20]. Showing only a thumbnail without text can lead to users underestimating the relevance of a page, while showing only text often leads to overestimation. Combining both approaches by placing text below the thumbnail achieved the best recognition results [5]. However, thumbnails have to have at least a size of 208^2 pixels to achieve a recognition rate of above 80%, which limits their usage [40]. While thumbnails perform well for recognition, they provide nearly no benefits when users have not visited the page before [14]. Teevan et al. looked at the best way to create

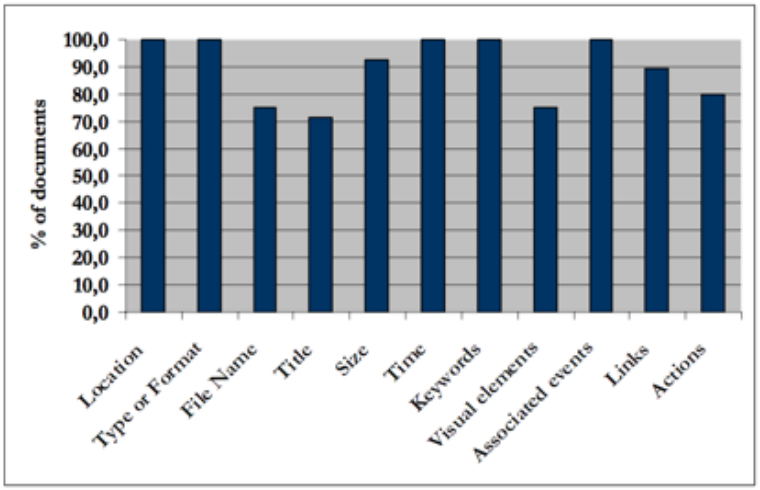

Figure 2.9: Searching for files on a computer is often a similar process to searching for information on the Internet. People have good memories about contextual information, such as how they got there (path), time of visit and also associated events [7].

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

Figure 2.10: Users generally remember visual elements and temporal information, while title and URL are rarely recalled [54].

recognizable images from websites and found that there are better ways than creating a miniaturized version of the page. By combining a page's logo, most prominent image and the title, they achieved a higher recognition rate while also using a smaller image [45]. As this approach is difficult to do automatically, using zoomed in thumbnails provides an easier alternative with a high recognition rate [5].

Besides visual elements, i.e., the structure of a page or prominent images, users also remember context like time, associated events or pages visited around the same time (see Figure 2.9 and 2.10) [7, 13]. The remembered time of visit can sometimes be skewed, because more memories about an event let it seem closer in time [8]. Therefore, users are generally not good at estimating the exact day of a page visit, but usually remember a broader time frame, such as "two weeks ago". As browsing is often a task-centered process, users also tend to have a good memory of websites and pages that were visited in sequence or around the same time, e.g., different results for the same query [18, 29]. Current browsers only provide a very targeted history search, which means it is only possible to search for specific pages, but not their context. If a user wants to rediscover a search result but does not remember the title, it would not help him if he remembered other results from the same session. Also, users generally do not remember everything about their browsing sessions when they start to rediscover a web page. They may remember separate pieces of information, like keywords of the targeted page or the rough query. Searching for keywords is often ineffective, as search engines do not customize results based on the user's history. Personalizing the user's search results has proven to be an effective way to increase search result quality [23, 35]. Further memories are likely to be triggered while the user is trying to find the desired page [25]. An example for this is a user trying to rediscover a search result by re-creating the query. At first, he may not remember all search results he looked

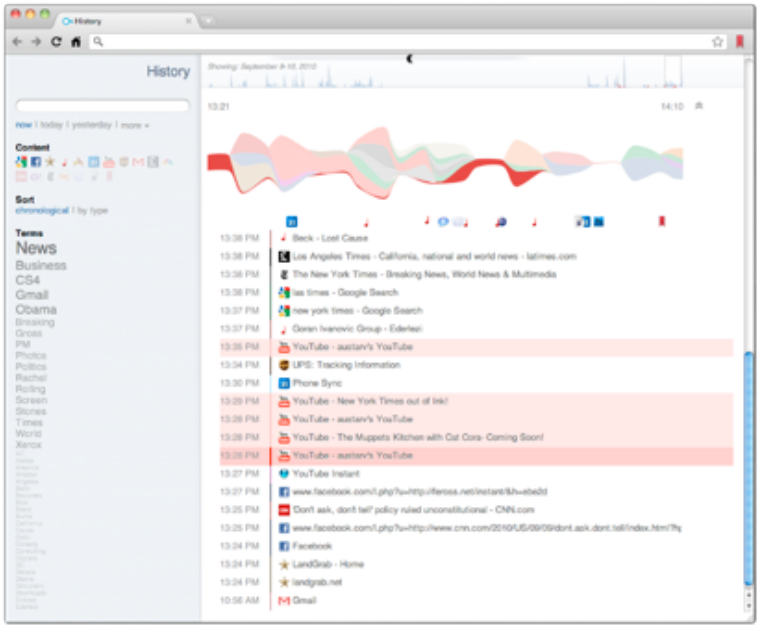

Figure 2.11: YouPivot combines the user's browsing history with contextual information, such as open applications, edited files or played music [18]

at and the exact path he took. When looking at the search result page, memories about visited results and subsequent visits might emerge. This process cannot be triggered by the browser's history list, because an unordered list that the user has never seen before is unlikely to trigger memories [25].

All in all, today's browsers provide little support for users' memory. As bookmarks are rarely used, the most suited browser tools are the history list or the address bar. Both tools require the user to enter a query which then acts as a filter for the users history. Users rarely remember the exact title or URL, but rather time, context and visual elements. Because these memories cannot be used, they fall back to inefficient and time-consuming strategies, such as re-tracing and re-searching.

2.4 Proposed Solutions

As long-term revisitation is often time-consuming, error-prone and poorly supported by browsers (see sections 2.3.1 and 2.3.2), many approaches and solutions have been explored. This section will give a brief overview of some approaches and how this thesis fits into these.

2.4.1 YouPivot

Hailpern et al. utilized users' contextual memories to build a revisitation tool called YouPivot. They found that users often remember events that happened while browsing the Internet, e.g., phone calls, music, physical locations or other environmental factors [18]. A server that runs on the user's machine collects contextual data such as open applications, played music or edited files. A chrome extension combines this data with the browsing history to show a visualization of all user activities (see Figure 2.11). YouPivot provides a significant advantage over the traditional history tool and also achieved a higher satisfaction rate.

2.4.2 LiveThumbs

Leiva et al. explored how thumbnail animations can be used to enhance website recognition. Instead of showing static images, this system animates thumbnails and allows users

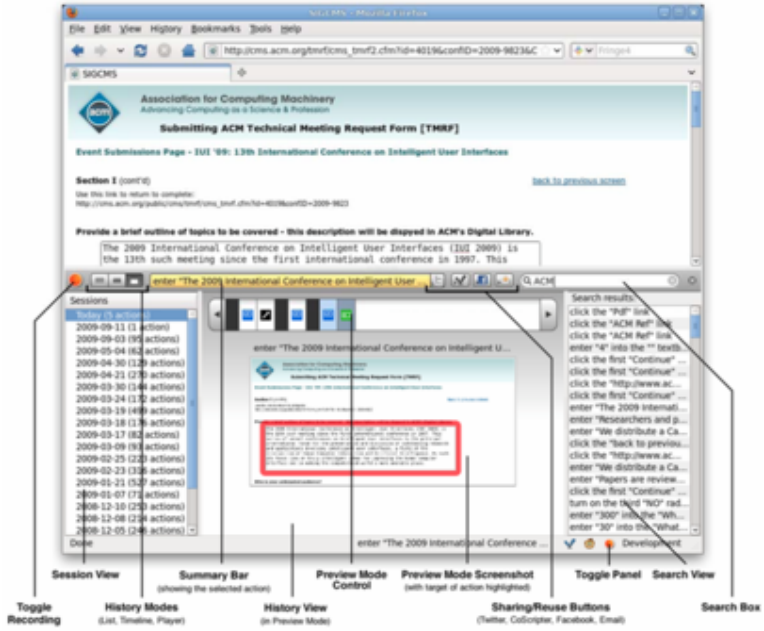

Figure 2.12: ActionShot shows detailed information about the user's history, such as mouse clicks, text inputs, browsing sessions and website screenshots. Users can share and replicate specific actions [28].

to look at other parts of a web page without having to visit it. They explored different animations like scrolling, zooming or highlighting specific parts. The evaluation showed that using LiveThumbs for recognition performs better than static thumbnails when users look for visual elements on a page. However, showing the animations also slowed the participants because they may had to wait for specific parts of the website to show up [27].

2.4.3 ActionShot

ActionShot is a browser tool that creates a more detailed view of the user's history. It shows fine-grained information about previous visits, e.g., buttons clicks, text inputs and website screenshots (see Figure 2.12) [28]. Users are also able to replicate specific actions or share what they have done on specific pages. The authors raised concerns about privacy and securely saving the browsing data, especially because there is no filtering of sensitive data. Overall, ActionShot provides information that is currently not accessible in browsers and allows users to better comprehend what they have done on the Internet.

2.4.4 SearchBar

Morris et al. identified that users often conduct complex search tasks that span over multiple sessions and involve various queries [36]. Returning to a search and re-finding specific results is an important part of these tasks. They created SearchBar, a system that stores the user's queries and makes it possible to save notes and ratings. The evaluation showed that users effectively used this interface to perform complex search tasks, including returning to previous queries and results.

2.4.5 Research Trails

Another approach by Liu et al. clusters the user's browsing sessions as visual streams [29]. By grouping websites that were visited around the same time, Research Trails provides an overview of the user's tasks and allows to resume them. Furthermore, users are able to inspect a specific session by looking at screenshots of the individual pages. Finally, the interface also contains meta information about visited pages, such as visit time, duration and keywords.

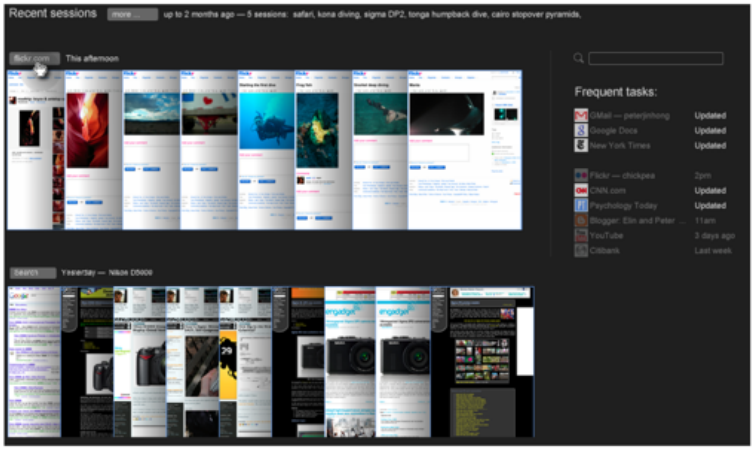

Figure 2.13: Research Trails shows the user's history as a visual stream and allows to resume previous browsing tasks [29].

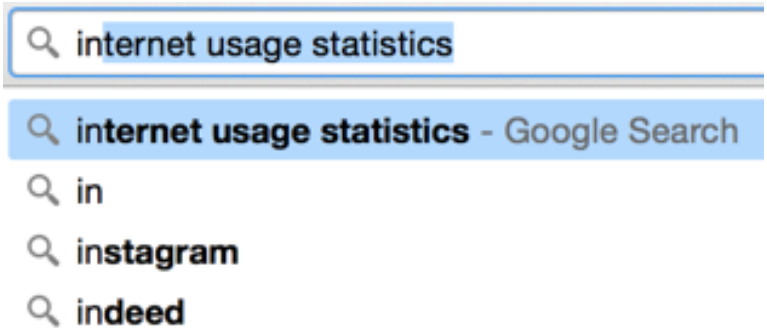

Figure 2.14: When the user input matches the beginning of a previous search query, Google Chrome autocompletes the query and opens the search result page.

2.4.6 Google Chrome Query Completion

Google Chrome offers another browser feature that helps users while re-searching information. While typing a query in the address bar, previous search queries will be autocompleted (see Figure 2.14). However, this only works if the user input matches the beginning of the query. When the user searched for "internet usage statistics", the autocompletion will be triggered by "internet", but not by "usage" or "statistics". This feature potentially saves the step of re-creating previous queries, but users still have to re-trace from the result page. When similar searches were done within a browsing session, the user might have to modify the query that was autocompleted.

2.5 This Work

There has been extensive research on the problems of long-term revisitation. It has become clear that rediscovering content and information is a problem that affects users and makes up a considerable proportion of all page visits (1 – 11%, depending on the user) [37]. As the tools provided by browsers see nearly no usage, many solutions have been explored (see section 2.4).

The presented approaches all focus on specific areas of revisitation. While YouPivot merges the browsing history with other activities on the computer, ActionShot provides a very detailed breakdown of all browsing actions, Research Trails visualizes sessions and SearchBar focuses on search tasks. All of these solutions perform well for different use cases:

- **YouPivot** supports complex tasks, where the user does different things in parallel
- **Research Trails** works best for getting an overview of task-based sessions and resuming them

- **SearchBar** is an easy way to view and return to previous results and queries
- **ActionTool** makes it possible to get a very detailed view of user actions and repeat or share them

This thesis aims for a more generalized solution. Instead of focusing on specific areas of revisitation and rediscovery, the goal is to create a history interface that makes it easy to rediscover content using a range of strategies. As the result is potentially integrated into the CLIQZ browser extension, this work will also put a strong focus on usability and interaction design. Therefore, general user behavior will exert a strong influence on design decisions. The final approach is not supposed to perform better for special use cases where solutions have already been explored. Instead, I will try to built upon the integrated browser tools to create a tool that is easily usable and makes it simpler to rediscover web pages by supporting user strategies, behaviors and memories.

3 Understanding Revisitation Behavior

Many studies have looked at revisitation and especially rediscovery and they all come to consistent conclusions: Rediscovering web content makes up a significant proportion of all page visits and is important to users. Furthermore, rediscovery is often a time-consuming process that is poorly supported by the browser and often fails (see chapter 2).

The way people browse the Internet constantly changes: Smartphones, tablets and interactive web content have all changed web interaction. Internet browsers are in rapid development and new features are added regularly. However, much of the related work dates back five or more years. It is therefore questionable if all findings apply today and if people still experience the same problems. To verify the previous studies, I performed a quantitative and a qualitative user analysis. While the former focuses on how often people revisit websites, how much time and how many clicks they need, the latter focuses on problems, strategies and user needs.

3.1 Study 1

3.1.1 Study Design

To gather quantitative data about website revisitation, I collected data from 26 users of the CLIQZ Firefox extension. All of these users work for CLIQZ and are therefore advanced users with higher than average computer skills.

To learn more about how these users revisit websites, the following usage data was recorded by analyzing their browsing history:

1. The proportion of web pages that were visited exactly n-times. As rediscovery is mostly done for pages that were only visited once before, it is most important to look at pages that were visited exactly twice (To exclude the back button and navigational links, revisits inside a single browsing session are not counted).

2. The average time between visits of the same page. This time helps to understand in what time frames users revisit and rediscover websites.

3. The time between the start of a browsing session and the visit of a specific page. By comparing this time to later revisits of the same page, an estimate of the revisitation performance can be identified.

4. The number of page visits from the start of a session to a specific page. By comparing this number for the initial visit to a subsequent visit, it is possible to infer that a shortcut was used (bookmarks, direct access, etc.).

The collected data will provide insights on how often users revisit pages, how they perform and what strategies they use. I hypothesize that rediscoveries make up a considerable amount of all page visits, take about the same time as the initial visit and that user's mostly use re-tracing and re-searching as their strategies.

3.1.2 Results

Proportion of Revisitation and Rediscovery
First, it is important to understand what proportion of all page visits are revisits. To classify as a rediscovery, a specific web page is usually only visited one time before a user

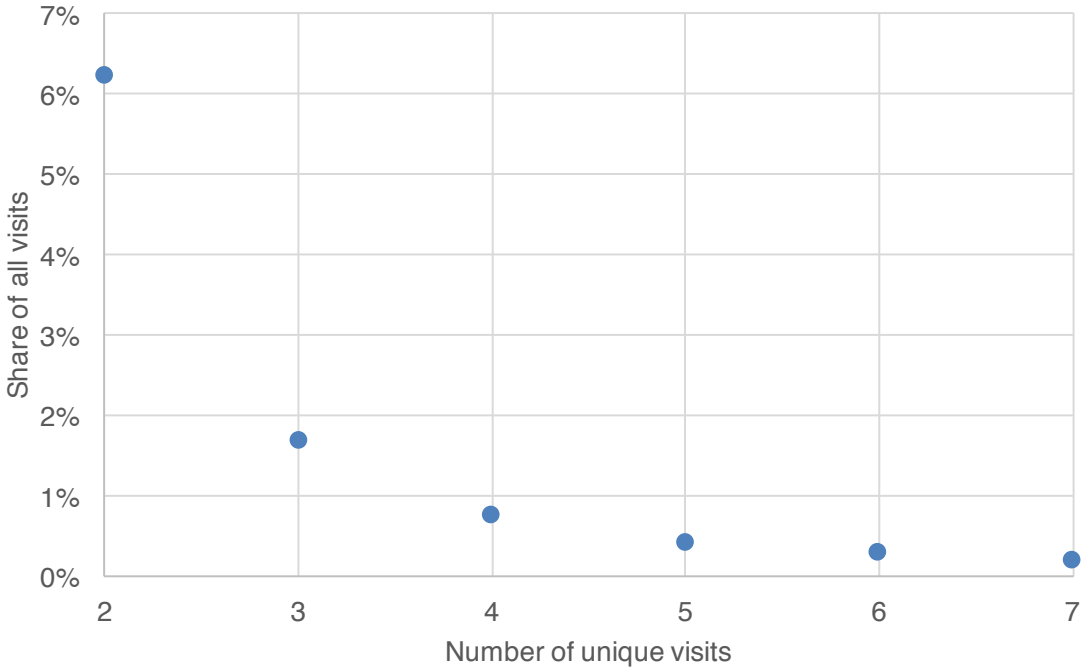

Figure 3.1: Proportion of pages that were visited exactly n-times. Most rediscoveries are URLs that were visited exactly twice: The initial visit and one revisit.

tries to re-find it. The more often a specific page is visited, the less effort it takes to visit it again. Therefore, it can be assumed that most rediscoveries are pages that the user has visited exactly two or three times, which means it was visited once or twice before the rediscovery. Visits within a browsing session (back button and navigational links) were excluded in this data.

The data shows that 89.6% of URLs from users' browsing histories are visited exactly once. Pages that were visited twice make up 6.2% and pages with three visits 1.6% (see Figure 3.1) which results in a rediscovery rate of up to 7.8%. By factoring in the average time between these visits, the estimated rate is about 3.05% (see Time between Visits). This rate does not include failed attempts because they cannot be observed. This data coincides with previous research which observed a rediscovery rate of 1 to 11%, depending on the user [37].

Time between Visits

Another important factor of rediscovery is the time that has passed since the initial visit of the desired page. This time helps to identify pages that will potentially be rediscovered and also influences the strategy used by users. For shorter rediscoveries, users might have better memories of titles and queries and prefer re-searching. Longer revisitation times might involve other strategies, such as searching for keywords from the targeted page.

To determine the revisitation time, I looked at the visit time difference between URLs that were visited exactly twice by users. The data shows that most revisits happen within 43 minutes and 7.2 days (median 19.5 hours), but users might revisit a page up to 28 days later. 49.2% of pages visited twice have a revisitation time longer than one day (3.05% of all visits) and 25.7% longer than a week (1.59% of all visits). I also compared pages that were visited exactly three and four times by calculating the average revisitation time of all subsequent visits (see Figure 3.2). URLs that were visited three times have a median revisitation time of 4.2 days and URLs visited four times are usually visited after 3.96

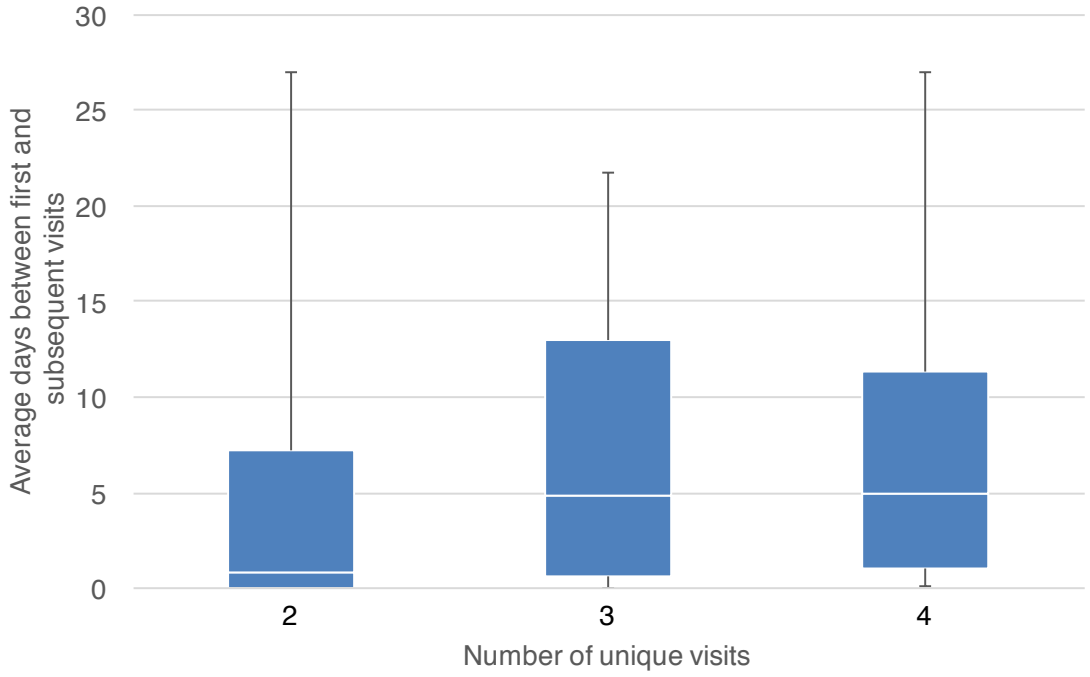

Figure 3.2: Time between visits of the same URL. While most rediscoveries, i.e., URLs that were visited twice, happen within 1 to 7 days, some users revisit pages up to four weeks later.

days. These times indicate that these pages are visited regularly over longer time spans and therefore include fewer rediscoveries than URLs visited twice.

The calculated numbers fall in line with previous studies from 2007 and 2008 [2, 37] which classified rediscoveries as revisits with a revisitation time longer than one day and often longer than a week. Using these thresholds results in a rediscovery rate of 3.05% for revisit times longer than a day and 1.59% for longer than a week.

Revisitation Performance (Time)

A major benchmark for revisitation is the time users need to find a specific page they have visited before. By comparing this time to the initial visit, it is possible to infer the user's strategy: A similar time suggests that the user re-searched or re-traced, i.e., he re-created his original path, while a significantly faster revisit signals a direct access or usage of the browsing history. This time difference also gives a good indicator of users general revisitation performance. In theory, the revisitation time should be faster because users have more information and memories about the page and are also able to use browser tools (history list, autocompletion, query completion).

To determine the revisitation performance, I compared the time from session start to the target page for the initial visit and the last revisit. The start of a session is triggered by different user events: address bar input, search bar input, bookmark actions and history actions. In other words, a new session is started when the user intentionally goes to a new website. Lastly, I subtracted the revisit time from the initial visit time, i.e., a positive value indicates a slower revisitation time and a negative value means the user was faster when revisiting (see Figure 3.3).

The data shows that users are generally about as fast when revisiting a page as when they initially visited it. These findings apply for URLs that were visited twice, i.e., mostly rediscoveries (median = -1.33 seconds), but also for pages that were visited three (median

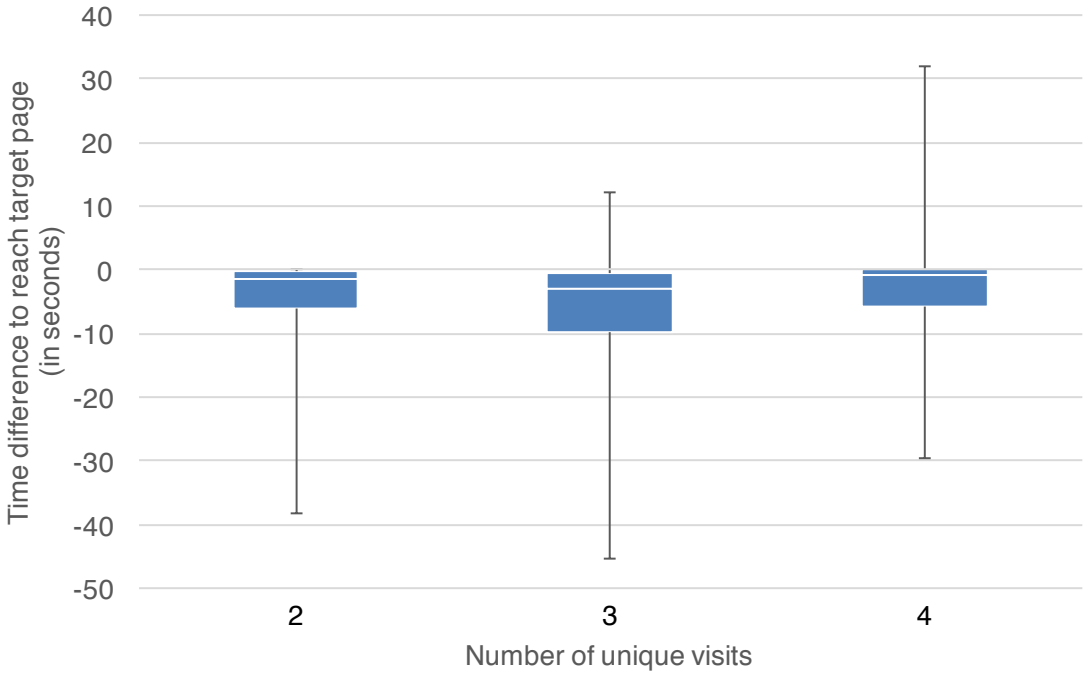

Figure 3.3: Time difference from session start to reach target page, comparing the first and last visit. The median time is close to zero for all cases, which means that revisitation time is mostly very close to the time needed to find the page.

= -2.66 seconds) and four times (median = -0.69 seconds). For rediscoveries ($n = 2$), users were rarely slower than the initial visit, but sometimes were significantly faster, which indicates direct accesses by using the history or bookmarks.

These findings coincide with previous research, which observed that the time needed to rediscover content is about the same as initially finding it [37, 39].

Revisitation Performance (Clicks)

To further analyze the performance of rediscoveries, it is also relevant to look at how many pages users visit when rediscovering content compared to the original visit. If users generally need the same number of page visits to revisit a page as they needed to find it, it can be assumed that they re-searched or re-traced the page. On the other hand, a lower page count for revisitation indicates the use of shortcuts or direct access.

The median difference between the number of visited pages for the initial visit and the revisit is zero for all URL counts (see Figure 3.4). Rediscoveries ($n = 2$) and more frequent revisits ($n = 3$, $n = 4$) are very similar in performance which suggests that users fall back to the strategy of their initial visit, no matter how much effort it takes. In other words, users tend to use a familiar browsing path rather than using other strategies that might allow them to directly access the content. These findings confirm previous research by Barrett et al. [6], who found that people like to use a known starting point when looking for information.

3.1.3 Conclusion

All findings that were presented comply with previous research. Although many studies range back five or more years, their discovered problems and user behaviors still apply today. I confirmed that users often go back to pages that they have visited only once

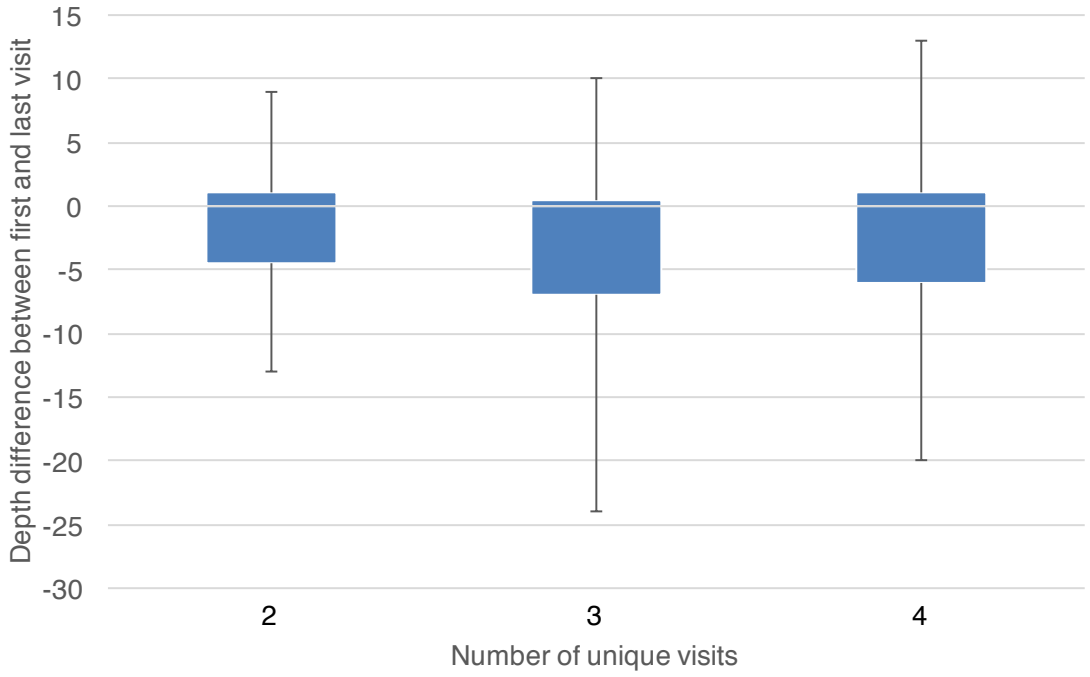

Figure 3.4: Difference between the number of visited links between first and last visit. The median link difference is zero for all cases, in other words users generally need the same number of page visits to revisit a page as they needed to first visit it.

before, sometimes up to four weeks later. The estimated rediscovery rate of around 3% falls within the predicted range [37]. Furthermore, the performance when rediscovering pages is almost identical to the initial visit. Users need about the same time and the same number of clicks to visit a page again. These results confirm that rediscovery is still a common task that involves re-creating the previous browsing path, i.e., re-searching or re-tracing the desired content. Finally, this data also shows that browser tools designed for revisitation are rarely used and there is still room for improvement to make rediscovery easier and more efficient.

3.2 Study 2

3.2.1 Study Design

To confirm the findings from the quantitative analysis and related work, I also conducted interviews with four CLIQZ employees. The participants had different backgrounds (marketing, search quality and project management) and stated that they have average or slightly above average computer skills. The interviews were split into two parts: (1) general questions about revisitation and (2) rediscovery tasks.

The first part of the interviews focused on the users' revisitation behavior and strategies. The following questions were used as a guideline and are based on the most important aspects of rediscovery that were identified in related work, i.e., frequency, importance, memories, strategies, problems and performance:

1. Do you often find yourself trying to rediscover websites?
2. When have you last tried to rediscover a website?
   (a) What was it and how important was it to you?

   (b) What did you remember when searching?
   (c) What was your strategy?
   (d) Did you have any problems?
   (e) Were you successful?

3. What browser tools do you generally use for rediscovery?

4. Compared to the initial visit, do you have the feeling that you are more efficient when re-finding visited pages?

In the second part of the interviews, the participants had to rediscover pages using their normal browsing environment. All pages were hand-picked from the their histories and represented common rediscoveries, ranging back at least one day, up to four weeks. The number of tasks per user ranged from 13 to 17. To get a comprehensive overview of the users' behavior, the tasks included a variety of different pages, e.g., search results, news articles, blog entries or videos. To simulate different scenarios, pages were presented in different ways:

1. Description of content

2. Screenshot of page

3. Prominent image or video from page

4. Description of a query that led to a specific result

As users' memories vary greatly for each visited page, these representations give a more comprehensive overview of strategies and behaviors. In one case a user might remember keywords or images from the page, while in another case he might think about context or his browsing path. All participants used Firefox and had access to the CLIQZ extension, but were allowed to use any strategy or tool of their choice.

I hypothesized that users still have the same problems that were reported in the related work and prominently use re-searching, re-tracing and keyword searching as their strategies. Lastly, the failure rate is hypothetically at least 7% [37].

3.2.2 Findings

Questions
All participants regularly rediscover pages from their browsing history, mostly within two weeks, but sometimes ranging up to a month. Furthermore, the sought information is generally important and justifies a considerable time investment.

The most common strategy, used by all participants, is re-searching the desired information. A shared problem is the re-creation of previous queries, which often requires substantial effort. Two participants also mentioned changing page content as a problem, especially on search result pages. Even when re-creating the exact query, the desired result might have disappeared or might be at a different position. Another problem occurred when several search results were accessed in sequence. In this case, users reported that they often have trouble identifying the correct result and have to look at all results again.

When rediscovering content that was not accessed by a search engine, the most common strategy is to search for keywords from the page. Re-tracing the content was mentioned as an less effective alternative.

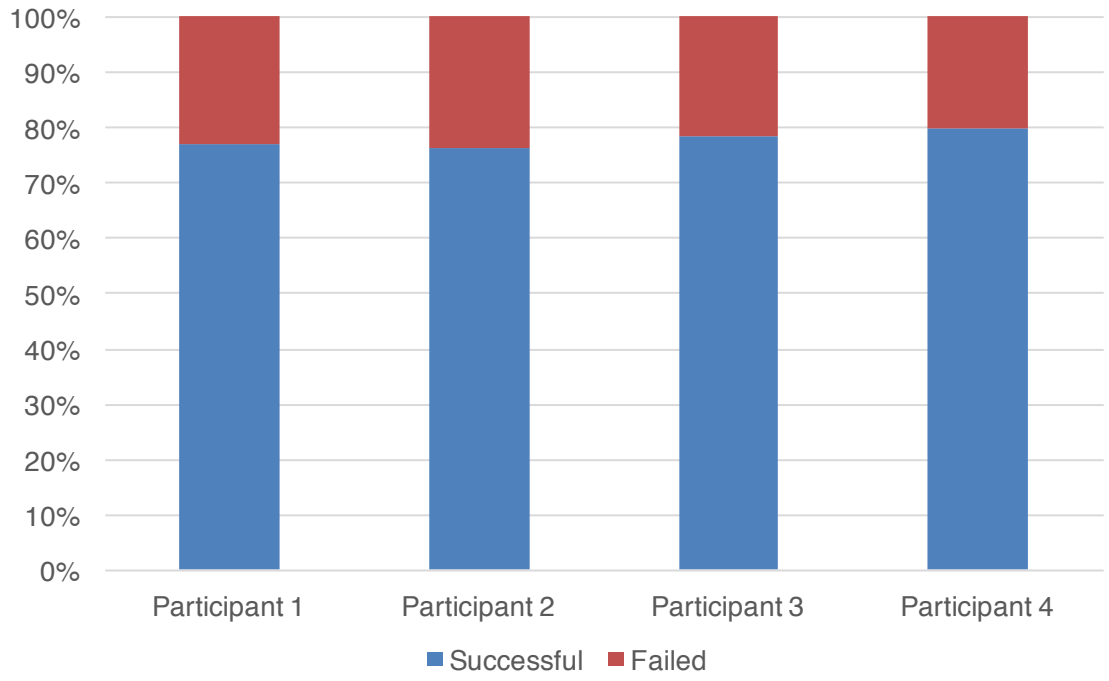

Figure 3.5: Success rate for the rediscovery tasks

Only one participant feels that he is faster when revisiting pages from his history, while the others need about as long or sometimes longer compared to the initial visit. However, everyone mentioned that they rarely fail and generally succeed when re-finding content.

The participants rarely use browser tools for rediscovery. The most commonly used browser feature is searching the history by using the address bar. As this approach requires the title of the page, users reported that they often fail or need a long time using this strategy. The browser's history list was not used or unknown. One participant described it as visually displeasing and not user-friendly: "I mean, just look at it, this is just no fun", while another was surprised that this feature exists: "Crazy! This really shows my complete history? Are you serious?".

Rediscovery Tasks

The success rate when users rediscovered content from their browsing history ranged from 76.5% to 80%, averaging at 78% (see Figure 3.5). The participants used a variety of strategies which were influenced by their memory and by the information that they were given about the targeted page.

When given a result from a previous search, the most used strategy was to re-search, i.e., re-accessing the result by searching for the same query. This approach often required several attempts because the participants rarely remembered what they searched for exactly. When looking at the result page, the visited ("purple") links often provided a big help to recognize the correct page, one user described re-search as "looking for the purple color". There was one case where a user did not remember the exact path that led him to the page, but his memories were triggered when he saw the highlighted results. After seeing the result page, he remembered the complex path of several links that lead him to the target. Another problem occurred when a high number of search results was viewed within a session. The participants had difficulties identifying the correct result and visited all purple links until they found the right page. In these cases, some users failed, especially when their initial session included several queries or the result page has changed since the

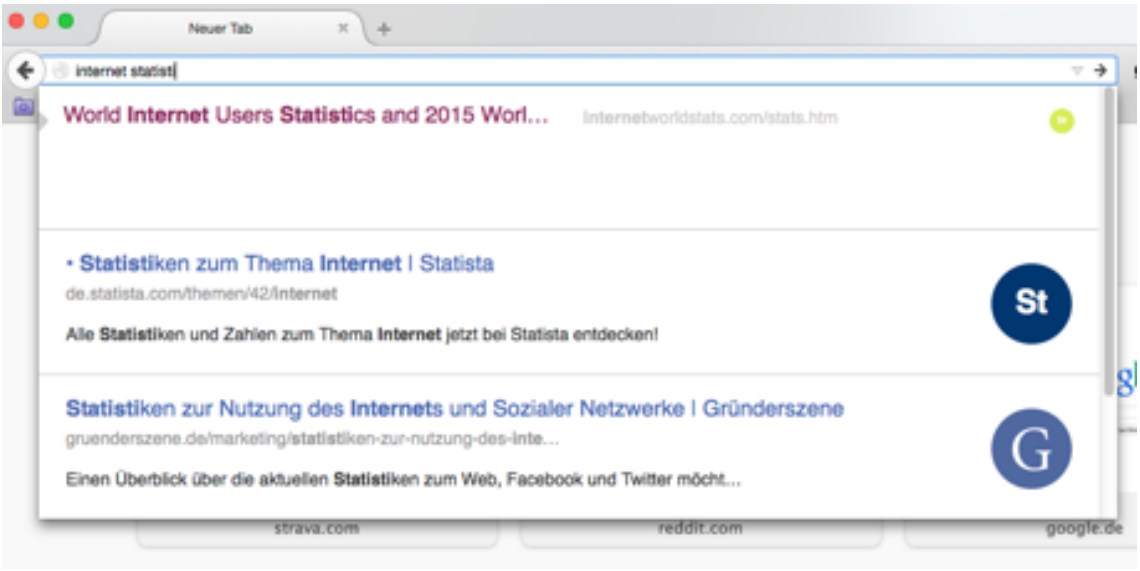

Figure 3.6: Users often had trouble identifying text links from their history. Although these links may lead to the target, they often preferred re-searching, which results in a longer, but known, browsing path.

last search. All in all, re-searching was a complex and time-consuming process, which one user described as a "completely new effort".

One common problem among users was the recognition of the correct page. In a few cases, the browser's address bar was used to start a Google query and the correct page was already shown in the dropdown. However, users were often unable to recognize this text link as the page they are looking for. As they usually did not remember the exact page title, they were not sure whether this link lead to their target. Generally, participants preferred a longer, but known, browsing path over a potential shortcut (see Figure 3.6). When following a longer path, users were also more certain about the page being the correct target.

In some tasks users where presented with a page that was not accessed by search, e.g., a news article or video. The most common strategy to access this content was to search for keywords on Google. This approach had a high failure rate, especially for videos and usually was only successful when users remembered the title of the article or video. Searching for keywords from the content was often difficult because the users' memories were not always comprehensive which lead to inaccurate queries and results. When participants remembered the domain, they sometimes tried to re-trace their browsing path. This approach often failed because the content had changed or they did not remember the correct domain or path.

A problem that was independent of the task and was shared among all participants was the inability to make use of contextual memory. In many cases, users remembered the date of visit, what they did simultaneously, what they visited before or even how they felt ("I don't remember anything except that I was extremely upset when I visited this page"). The tools that were used did not support any of these memories. A successful rediscovery often depended on the knowledge of the query and page title. If a user remembered neither, it took great effort to re-find the page. The browser's history list was used as a last resort in two cases, but did not yield any results. Users found it too difficult and tedious to look through the long and cluttered list. Using the browser's address bar to search the history was utilized by some users and succeeded when they remembered the title of the targeted page.

3.2.3 Conclusion

The findings from the qualitative study coincide with previous research and revealed the same problems, behaviors and strategies. All participants regularly rediscover content, mostly within two weeks. While it is generally important for them to find this content, it is often time-consuming and requires great effort. Almost all rediscoveries involved using a search engine, either in form of re-searching or using keywords from the sought page. Users preferred longer browsing paths over shortcuts and often used known starting points for their rediscovery. Around 22% of rediscoveries failed, often because users did not remember the query or page title and could not make use of contextual memories.

3.3 Summary

Both studies confirm the findings from previous work. Rediscoveries are still a common task and are often time-consuming and prone to failure. Compared to the initial visit of a page, users generally need about the same time to rediscover it. The most common strategies are re-searching, re-tracing and keyword searching. These strategies are mostly unsupported by browsers and users rely on search engines instead. Recognizing the correct page was a common problem and users rather used a longer, but known, path instead of using shortcuts.

When users rediscovered pages from their history, they often had trouble remembering enough information that they can use to initiate a search. Browser tools require knowledge of the URL and title, while search engines need precise information because they do not prioritize the user's browsing history. Therefore, it can be assumed that rediscovery performance can be greatly increased when users are offered a tool that assists them in common strategies and provides support for contextual memories. These insights lead to the design of *CLIQZ Browsing History*, a history interface that acts as an alternative to the browser's history list and offers support for the identified behaviors and memories.

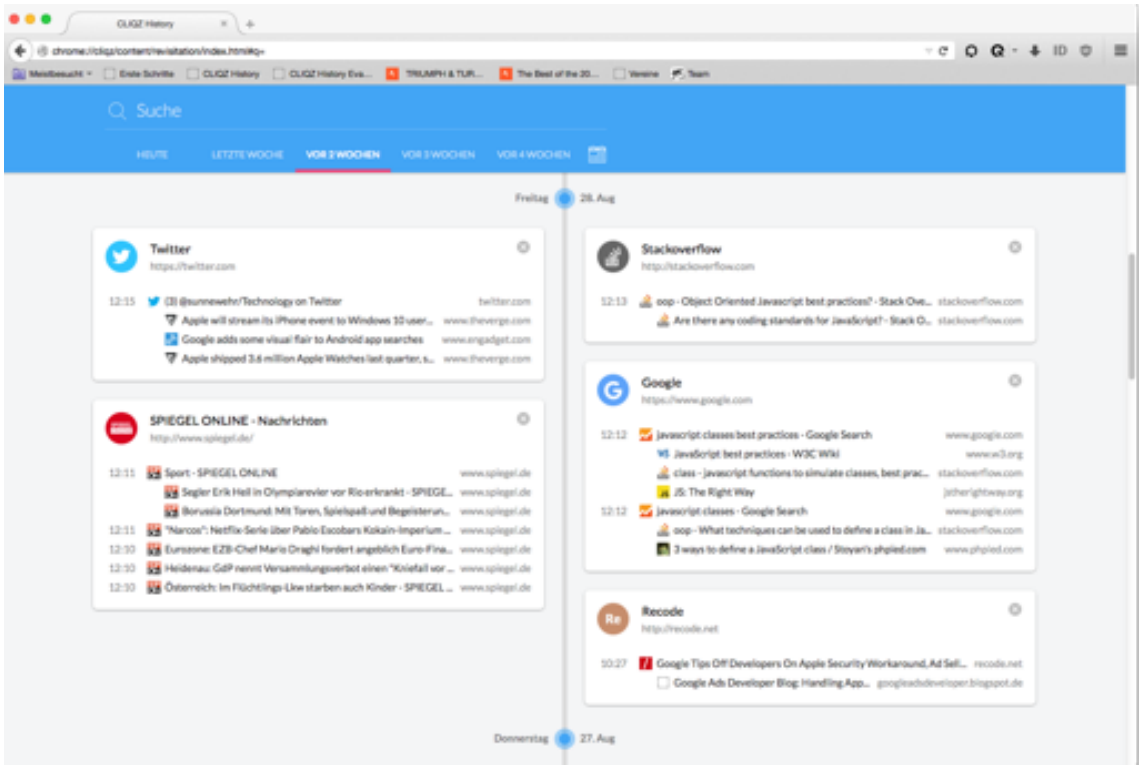

Figure 4.1: The main view of the CLIQZ Browsing History groups the user's history into sessions and makes it possible to comprehend browsing paths.

4 CLIQZ Browsing History Prototype

4.1 Overview

I developed *CLIQZ Browsing History*, a history interface that acts as an alternative to the browser's history list. Revisitation and rediscovery are made easier by providing features that support the identified user strategies and behaviors.

The main view of the developed tool shows an overview of the user's browsing history in form of a timeline, with the most recent sessions being shown at the top (see Figure 4.1). Each browsing session is represented by a separate card which shows visits that originate from the same domain. For example, when a user searches on Google and visits different results, these visits are all grouped within the same card. Another example could be reading different articles on a news website and visiting links that lead to external websites. Visits within a session are indented to represent the hierarchy and browsing path. All in all, this view makes it easy for users to identify and comprehend their history. The visibility of browsing paths and session starting points immediately gives an overview of browsing sessions and supports contextual memories.

The interface uses infinite scrolling and loads additional sessions on demand, i.e., the user can navigate his complete history by scrolling in the main view. There are five buttons on top which make it possible to automatically scroll to specific points in time. These buttons represent broad time frames (one to four weeks) which reflect common rediscovery times and support users' fuzzy memory of visit dates. There is also a calendar button which can be used to jump to specific days. As the whole history is shown in a

Figure 4.2: The search function suggests previous queries, even if it only partially matches the input

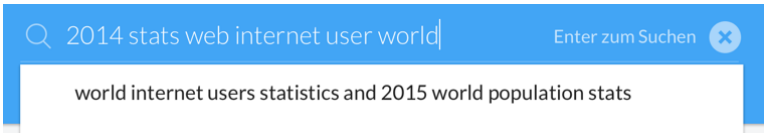

Figure 4.3: Entering a high number of keywords improves the quality of suggestions

single view, it is still possible to scroll to previous or following sessions after using one of the shortcuts.

Furthermore, there is a search function which offers suggestions to support users' memories (see Figure 4.2). These suggestions consist of previous queries and page titles, which do not have to match the query exactly. Even if words only partially match or are in the wrong order, there will still be relevant results. The search will then act as a filter for all sessions, i.e., all sessions that contain the query in a URL or title will be shown. To always provide context and support contextual memories, visits inside sessions are not filtered. For example, when a user selects a previous query as a search suggestion, he will still be able to see his whole search session, including all related queries and visited results. Lastly, all words that match parts of the query are highlighted inside the session, which makes it easy to identify the most relevant visits. Overall, the search feature is designed to make re-searching easier and more efficient. Users do not have to re-create their query, instead they can search through their queries and directly choose the correct one. Afterwards they are provided with all visited results, related queries and see their browsing paths. This feature can also be used to search for results that were not accessed by search engines which often involves searching for keywords. The search will ignore words that are not found and still show suggestions. Therefore, the search supports a very common strategy that is often used in this context, i.e., adding additional keywords to refine the search (see Figure 4.3).

Another common revisitation problem is the recognition of the correct page. As users are generally not good at recognizing titles, CLIQZ Browsing History also shows visual and textual information on mouseover (see Figure 4.4). This preview contains the URL, title and thumbnail of the page. If available, Opengraph [15] data will be used to show the most prominent image and a description of the page. Furthermore, the browsing path that lead to the mouseovered page will be highlighted inside the session. Combining all of these pieces of information offers an extensive support of common memories, i.e., visual elements, page structure, content and browsing paths.

All in all, the *CLIQZ Browsing History* provides a comprehensible overview of the user's history and offers several features that directly support rediscovery strategies and common memories. By grouping the history into sessions and showing browsing paths, users are able to easily reconstruct their history. The search feature supports the most common strategies, i.e., re-searching and keyword searching. Finally, an extensive page preview makes it easy to identify pages without having to open them.

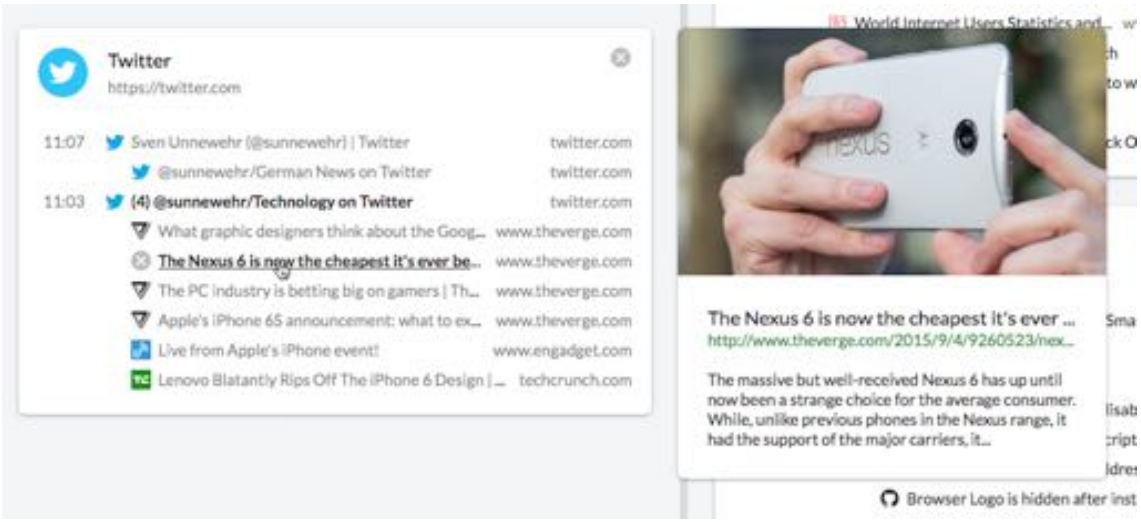

Figure 4.4: On mouseover, additional information about the page is shown. If Opengraph data is not available, an enlarged thumbnail of the page will be shown instead of the prominent image

4.2 Requirements

The related work and user analysis revealed several flaws in current browsers when it comes to long-term revisitation:

1. Tools that were designed for long-term revisitation see little usage (Bookmarks 3% [25] and history list 0% [9] to around 1% [4])
   (a) Managing bookmarks requires great effort and decision making
   (b) The history list is mostly unknown, has several usability problems and does not support common memories (see section 2.3)
2. Using the address bar to search the history requires memory of the title or URL
3. Re-searching is only partially supported and still requires high effort, such as re-creating queries, looking at visited results and re-creating the previous browsing path
4. Re-tracing is not supported by any browser feature

Re-searching and re-tracing are the most common strategies, but require high effort to successfully rediscover a page. Users need about the same time to re-find a page as they needed to find it, sometimes even longer. Furthermore, users often fall back to keyword search when they do not remember enough information about their initial session or were unsuccessful using the other strategies. However, this approach requires high mental effort and rarely succeeds. As these strategies are mostly unsupported by browsers, it can be assumed that rediscovery performance can be greatly increased when users are offered a tool that assists them in these behaviors. Therefore, there were several objectives when designing the prototype:

1. Support re-searching
2. Support re-tracing

3. Intelligent keyword search
4. Easy page recognition, e.g., the user should not have to visit all search results again to find the correct one
5. Support for contextual memories
6. Easy to use, little complexity

There have been different solutions to these problems in the past (see section 2.4). However, my intention is to develop a tool that does not provide a specialized solution to one of these problems, but rather a generalized approach that can be used for all use cases. The aim is to develop an easy-to-use tool that can be used as a stand-alone history list while also supporting the identified strategies. Such an all-in-one solution has the advantage that users can use it as a sole starting point for all rediscoveries. Combining all rediscovery strategies into one reduces the mental burden and makes re-finding content simpler overall.

4.3 Design

4.3.1 Interaction and Interface

To support all rediscovery strategies, there has to be an overview of the complete browsing history which allows to look through all previous visits. This is especially helpful when a user only remembers the broad time of visit and has no other specific memories that could be used to start a search. By looking at all of his visits, he can then locate the desired page. This functionality is currently provided by the browser in form of the history list. However, there are two major downsides to this list: First, it is cluttered and ordered by time. As users often do several tasks in parallel, this list mixes all visits across tasks. Because of this, it is not possible to comprehend previous browsing sessions and paths. Secondly, all visits are only represented by the title, which makes scanning and especially recognizing pages difficult. These two factors combined make usage of the history list time consuming and often unsuccessful. To solve these issues, a history overview has to structure all visits in a way that is comprehensible for users.

As browsing is a task-centered process [30], it makes sense to use this as a starting point for structuring the history. A task might consist of several browsing sessions, scattered over a longer timespan. However, grouping the history by tasks is problematic for different reasons: Users' memory usually does not contain a task as a continuous activity, especially when it spans across multiple days or weeks [3]. Furthermore, automatically detecting tasks is error-prone and could therefore cause failures when rediscovering pages. An easier and simpler solution is to group the history only by sessions. Detecting a single session is straightforward, because it is started by pre-defined events, such as going to an URL, starting a search or clicking on a bookmark. Sessions are still easy to recognize for users, because they represent an isolated activity with a single intent. For example, a session could be a search that is started on Google and then leads to different results. In the user's memory, this search is a single activity that he remembers as an entity and can recognize by seeing the query and visited results. Therefore, it is also easy for users to map these sessions to the task they are related to.

Besides grouping the user's history into sessions, it is also important to represent browsing paths. Contextual memory is very common and users are therefore very good at recognizing their previous paths, which creates a higher confidence when rediscovering [39]. As re-tracing and re-searching are the most common strategies for long-term revisitation [37], showing the browsing path is essential to support these behaviors.

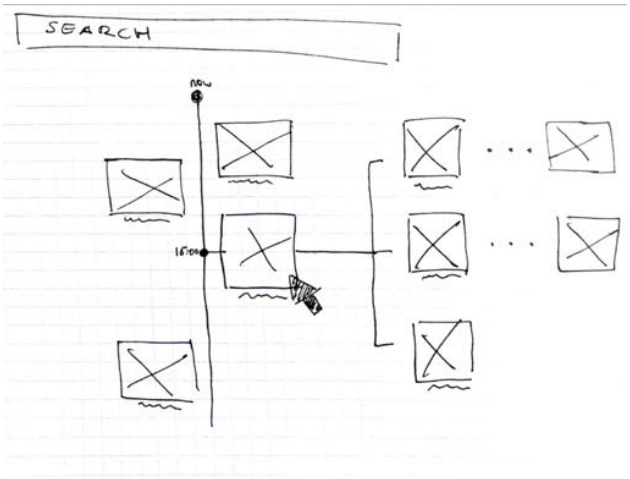

Figure 4.5: One of the initial sketches focused on an interactive approach, where browsing sessions could be navigated step by step. There is a timeline on the left where each session is represented by an image or thumbnail of the starting page. When clicking on these sessions, further visits inside these sessions can be navigated and specific paths can be re-created.

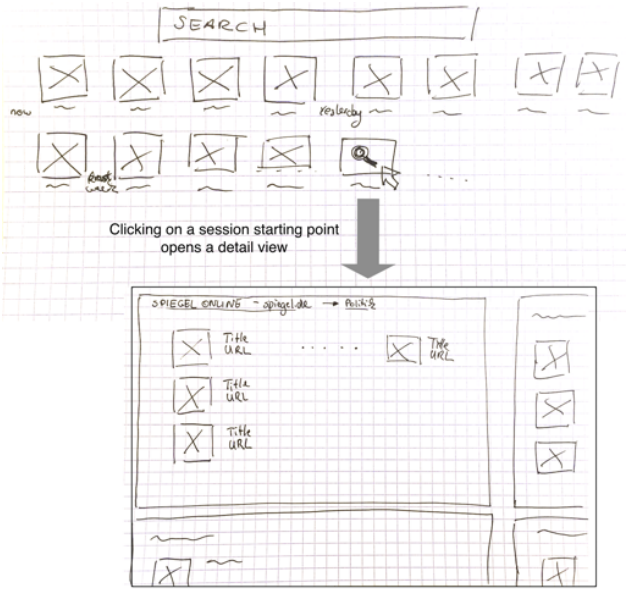

Figure 4.6: In another approach, each session's starting point is represented by a thumbnail (top part). When clicking on a thumbnail, the session will be zoomed in. In this detailed view, the user can navigate the session and re-create browsing paths. Furthermore, adjacent sessions can be accessed quickly by moving the zoomed section.

The presented insights lead to two major design decisions: (1) The browsing history should be structured into sessions and (2) browsing paths should be visible.

Grouping by sessions and showing browsing paths addresses two of the initial objectives, i.e., support for contextual memories and re-tracing. Page recognition is also made easier by showing the path, but can still be difficult, especially when comparing similar pages such as search results. Previous research has shown that thumbnails are an effective way to further improve the recognition of visited pages [40]. As thumbnails need a minimal size to have an effect, showing the complete history including sessions and paths is not feasible. Therefore, an initial approach to the prototype shows only parts of the history at a time, while the user can interactively navigate through sessions (see Figure 4.5). In this approach, session starting points are arranged in form of a timeline and represented by title, URL and a thumbnail of the page. When clicking on a session's starting page, pages that were visited afterwards are shown and previous browsing paths can be re-created step by step. As users are often interested in the last page of a specific path, these pages are shown at all times. I explored alternative interfaces which utilized the same interactive approach (see Figure 4.6).

Interactively navigating sessions offers several advantages when re-searching or re-tracing. The user is able to use the same strategy, i.e., following a path he remembers, by using a single interface. Instead of loading each page separately, he can identify pages by

Figure 4.7: An early functional prototype to test navigation inside sessions. For every page within a session, all links that were visited from that page are shown as thumbnails. The current path is shown above the thumbnails and the links can be used to navigate within the path.

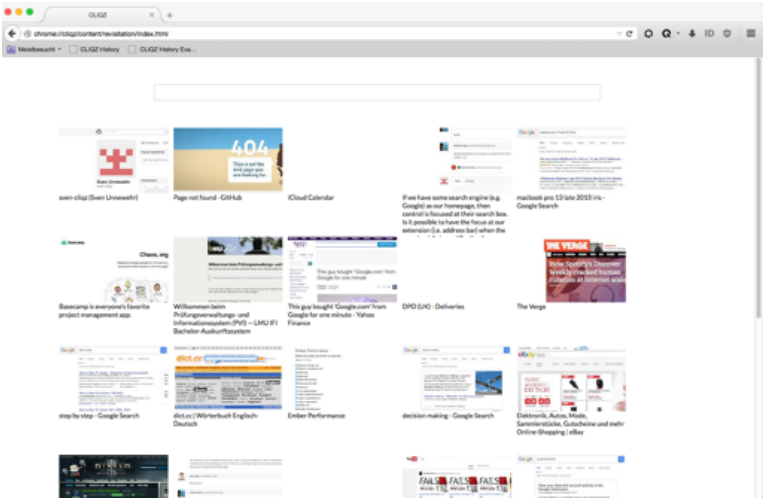

Figure 4.8: This prototype was created to get a feeling of the overall history view. Each browsing session is represented by a thumbnail of its starting page and the page title.

looking at the thumbnails and quickly navigate between pages. Furthermore, the presented interfaces only allow the user to choose a path that he already visited and therefore prevent mistakes. Navigating sessions replicates the previous browsing experience and allows faster page recognition and navigation.

To test the presented approach and compare different interfaces, I built a functional prototype (see Figure 4.7) which focuses on the navigation inside sessions. After selecting a session, each page that was visited consecutively is represented by a thumbnail and title. It is possible to re-create an arbitrary path from the session and also return to previous pages along the path.

Another prototype was developed to test the main interface, which shows the user's complete history and groups it into sessions (see Figure 4.8). Each session is represented by a thumbnail of its starting page and the title. All sessions are ordered by time, with the most recent session in the top left.

Both prototypes revealed several problems and flaws in the initial assumptions. On the one hand, grouping the history into sessions has the desired effects: The history is less cluttered, easier to comprehend and divided into coherent parts. Furthermore, the thumbnails made recognizing visited pages easier. On the other hand, only showing the starting page of a session is not sufficient. It is very difficult to identify sessions when only seeing the starting page, especially because it is often shared by other sessions. The user has to make a decision on which session to explore. If the desired page is not in the chosen session, the user has already put in high effort and has to start over. This problem is amplified by the interactive exploration; to be sure that a session does not contain a specific page, all potential paths have to be investigated. These problems lead

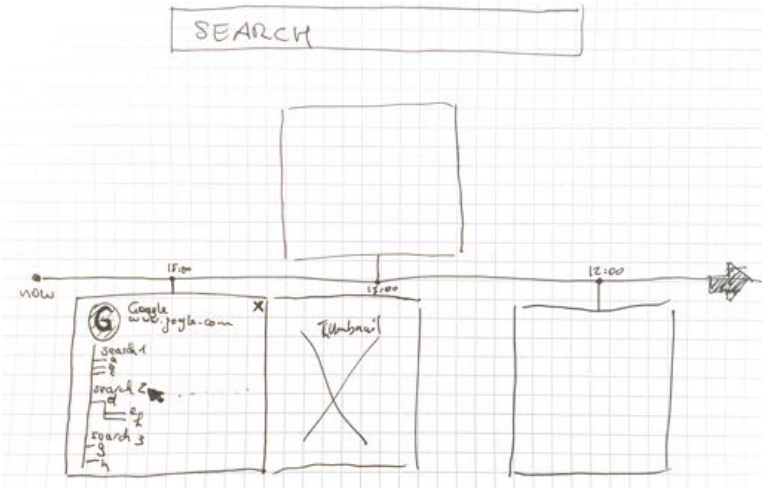

Figure 4.9: The final sketch arranges sessions along a horizontal timeline and shows all visits within a session as a list. The browsing paths are represented by the order and indention of each visit.

to the following insight: The main view has to provide a complete overview of all visits inside each session. By doing this, the history list can be scanned with little effort and the user does not have to make potentially penalizing decisions. Therefore, the approach of an interactive session navigation is also discarded. When already showing a complete overview of the session, interactive navigation does not provide any benefits. Additionally, the interactive prototype revealed usability problems when analyzed in isolation. Using a more complex interface raised the learning curve and giving the user a high number of decisions increased the mental effort and error potential.

Combining these insights with the previous findings leads to the following design guidelines:

1. The history has to be grouped into sessions and browsing paths have to be visible
2. The complete history, including all sessions with all visits have to be shown in a single view
3. There has to be minimal interaction, mental effort and potential for failure
4. The history has to be easily scannable
5. Thumbnails have to be shown in some way to improve recognition

Always showing thumbnails conflicts with other guidelines: As they need a minimal size to be effective [40], arranging them to show browsing paths needs a significant amount of screen space. This causes scanning to be difficult and leads to a cluttered and less comprehensible interface. Therefore, I decided not to show thumbnails at all times, but rather make them available on demand, e.g., by mouseover. This allows the design of a more compact, scannable and overall simpler interface, while still providing easier page recognition.

These insights lead to another draft (see Figure 4.9) which arranges sessions along a horizontal timeline and shows all visits inside a session as a list. This list represents browsing paths by intending visits according to their link depth and source page, comparable to tree views in file browsers. This concept is widely used, easy to grasp and provides a scannable overview of all visits in a session. Thumbnails of pages are shown on mouseover next to the visit list. Finally, the user is able to go back and forth in time by using buttons on the left and right edge.

When translating this concept into a prototype (see Figure 4.10), I decided to switch to a vertical timeline and to arrange the sessions from top to bottom. Vertical scrolling is

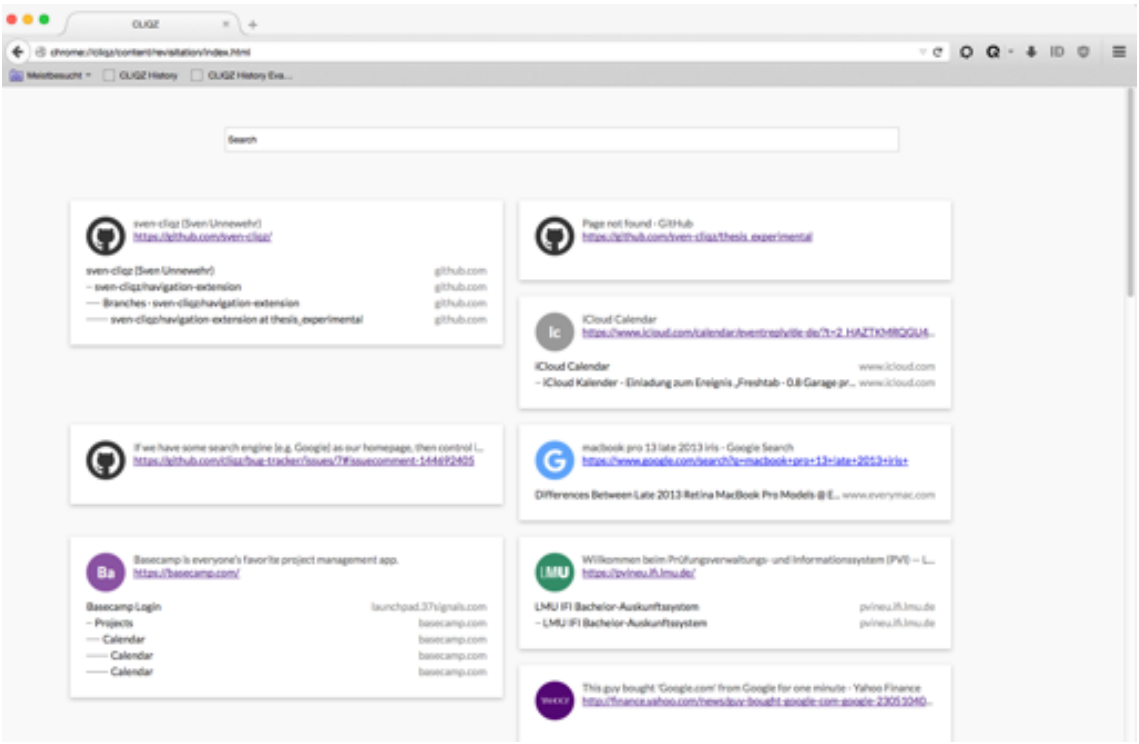

Figure 4.10: The final functional prototype introduces a scrollable list of previous browsing sessions. Each session contains a list of all its visits.

more common and natural and supports the usage of the mouse wheel instead of relying on buttons. Furthermore, this approach makes it possible to navigate seamlessly through the history as it does not rely on separate pages.

This prototype fulfills all guidelines and represents the foundation for further development. By grouping the history into sessions and listing all visits inside sessions, it is possible to easily scan and comprehend the complete browsing history. Browsing paths inside sessions are visible and reconstructible without taking much screen space. The interface requires minimal interaction; the user can go back in time by scrolling and thumbnails are shown on mouseover. This approach does not involve any additional decision making besides selecting the page to visit. The mental effort is reduced to a minimum, as thumbnails provide a way to identify pages before visiting them. All in all, the prototype meets all requirements and provides a simple interface that requires little effort to use. The next sections will go into detail on designing the interface and further features that address rediscovery strategies.

4.3.2 Sessions

Grouping the browsing history into sessions has shown to be an effective way to create a comprehensible structure. By ordering and intending the visits according to the user's browsing path, sessions are scannable and easy to grasp. This representation can be improved further by optimizing the visualization and structure.

Each browsing session has a specific start page, e.g., a news website that the user accessed by entering the URL or a search result page. When rediscovering, users tend to go back to these starting points and re-create their browsing path. To support this strategy and users' memory, each session prominently displays the start page, together with its logo and URL (see Figure 4.11). By doing this, users are quickly able to recognize

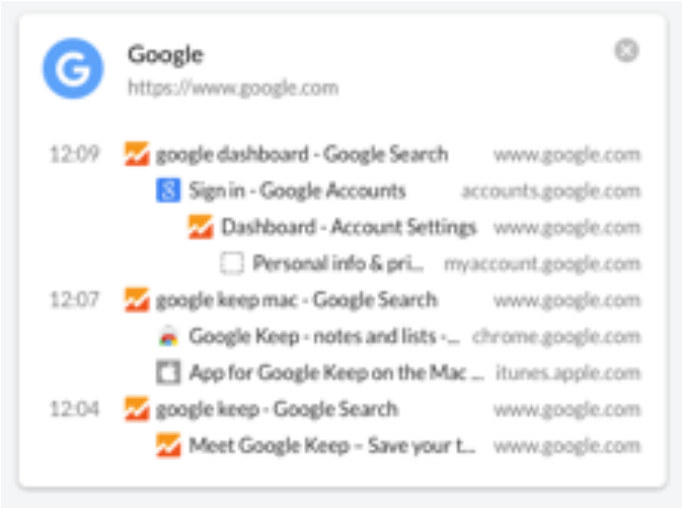

Figure 4.11: Each session prominently displays the start page, together with its logo and URL. When sessions origin from the same domain and happen within a certain time frame (30 minutes), they are automatically grouped. In this case, each search originally belonged to a separate session.

sessions and are supported in re-finding a previous path.

Sometimes, users do multiple browsing sessions within a short time frame, e.g., when sequentially searching on Google. Research has shown that sessions within a certain time frame (around 30 minutes) are very likely to be linked and are also connected in the user's memory [55]. To support this, browsing sessions that origin from the same page and happen within 30 minutes are automatically grouped. For example, when users are looking for specific information, they might need several searches and need to look at several results to find it. Although these searches could technically be separate sessions, it makes sense to group them and show them together (see Figure 4.11). This concept also applies for non-search sessions, e.g., when sessions are interrupted or resumed within a short time. A user might go to a news website, then decide to leave it to read his mail and shortly after return to the news site. Although these visits are different sessions, the user perceives them as one and mentally links them together.

To further improve sessions' scannability, visits are also represented by their favicon, domain and time of visit (see Figure 4.11). To prevent clutter, visit times are only shown for the first level of a session. By doing this, users get a general feeling of time frames and need less effort for scanning. Favicons and domains offer a quick way to recognize and distinguish pages [26]. Finally, the indentation makes browsing paths visible and shows which page was visited from where.

All in all, the presented way of displaying browsing sessions provides a quick and easy way to grasp and scan sessions. Users are able to comprehend visit times and browsing paths, while also being able to use favicons and domains to identify pages. Finally, users' memory is supported by grouping mentally connected sessions that happen within a short time frame.

4.3.3 Previews

The session representation supports page recognition by showing title, domain, path, favicon and time of visit. However, there are cases where this information is not sufficient, for example because a user visited many similar looking pages and only remembers visual elements or the structure of the page. In this case, research has shown that thumbnails provide a significant improvement for recognition [40]. For this reason, the prototype shows a thumbnail of the page on mouseover (see Figure 4.12).

In addition to the thumbnail, I took several measures to provide further preview information about visited pages. Many websites have started including Opengraph data [15] which includes additional information about the page such as a description or the promi-

Figure 4.12: To improve website recognition, a thumbnail of the page is shown on mouseover.

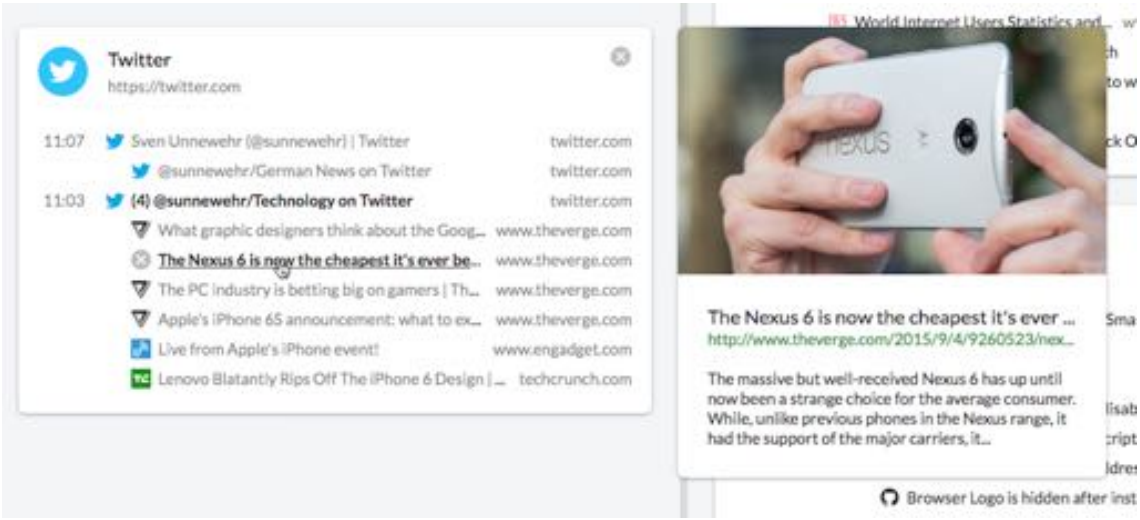

Figure 4.13: When Opengraph data is available, it will be used instead of the thumbnail.

nent image. When available, this data is used instead of the thumbnail (see Figure 4.13). As users often remember prominent images and keywords [54], this information greatly improves page recognition. Lastly, the browsing path that lead to the page is highlighted on mouseover.

In summary, the user is assisted in many ways when recognizing visited pages. By highlighting the browsing path, presenting visual hints and showing textual information, common memories are utilized and supported. The user is able to quickly compare different pages without having to load them separately. Finally, this feature takes little screen space and does not require the user to make potentially wrong decisions.

4.3.4 Timeline

Another important aspect of the interface is the arrangement of browsing sessions. The simplest approach is to order sessions according to their visit times which has different advantages in the context of the prototype. First, sessions close in time are often linked in users' memories and also might share similar purposes. By displaying them close to each other, additional memories can be triggered and comprehending past sessions becomes easier. Furthermore, ordering by time is intuitive and easy to understand. Alternative

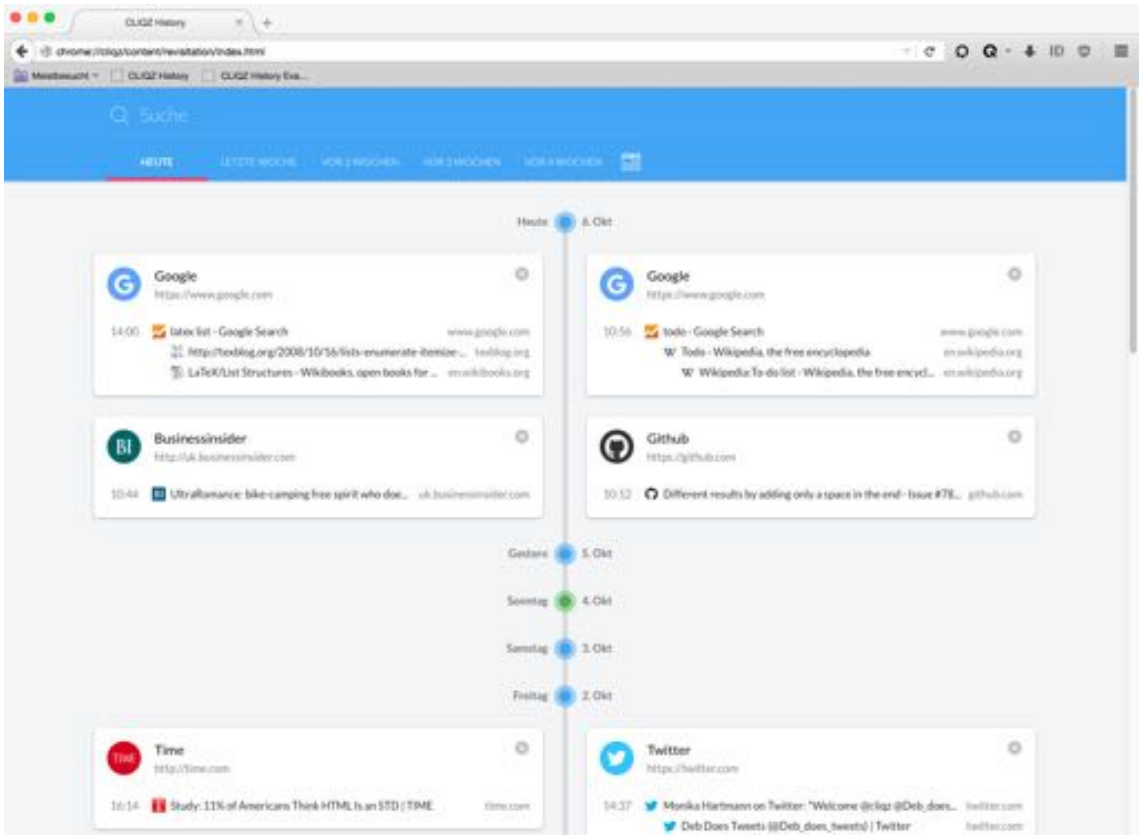

Figure 4.14: Sessions are arranged in two columns along a timeline. The history can be navigated seamlessly by scrolling up or down. Days are shown along the timeline to give the user a feeling of visit dates.

approaches, e.g., ordering sessions by domains, keywords or length, are difficult to grasp and require learning and understanding of the concept. These approaches are also more error-prone, whereas the visit time is a never-changing, definite value. Finally, navigating sessions by time is a common strategy when users use the visit date as a starting point for a rediscovery. All in all, ordering sessions by time is easy to understand, supports common revisitation strategies and is not prone to errors.

To make the user aware of the chronological order of sessions, they are arranged along a timeline from top to bottom (see Figure 4.14). By grouping sessions into two columns, a higher number of sessions is visible at once which creates a better overview and increases scannability. If they were displayed in one column, scannability would be decreased while not providing much more information. Users can go back in time by scrolling down, which causes additional sessions to be loaded on demand. This approach has the advantage that sessions are never removed from the interface and the user can navigate his complete history by scrolling up and down. As there are no separate pages, adjacent sessions are always visible and the history can be navigated seamlessly. Additionally, a scrollable, endless timeline is a well-known concept, e.g., used by Facebook or Twitter. To give the user a general feeling of visit dates, weekdays are shown along the timeline.

Users often remember the broad time frame of a page visit [54] and use this as a starting point for rediscovery. Therefore, I included several buttons in the prototype that can be used to jump to specific points in time (see Figure 4.15). The time frames of one to four weeks were chosen based on common rediscovery times and support the often fuzzy memories of visit dates. By doing this, users can use their memory as a starting

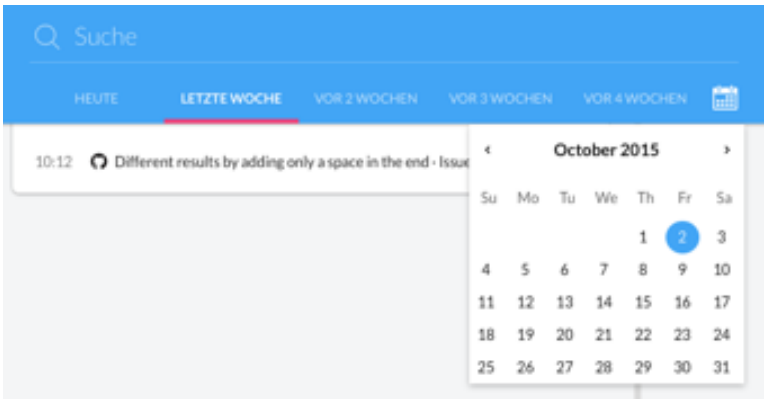

Figure 4.15: The interface includes buttons to jump to specific points in time. When manually scrolling, these buttons also are highlighted to indicate which segment of the history is currently shown. The calendar button can be used to directly scroll to specific days.

point and further investigate their history from there. When clicking on a button, the view is scrolled to that point, which also causes all intermediate sessions to be displayed. Choosing a wrong time frame has no negative consequences as the user can just scroll up and down from that point in time. To support cases where users have exact memories of visit dates, there is a calendar button which can be used to jump to specific days. Lastly, the buttons are automatically highlighted when the user manually scrolls and reaches these time frames. This provides a further indicator on which point in time is currently shown in the history.

In summary, the timeline provides an easy to grasp concept that makes it possible to easily navigate the history. By providing features to quickly scroll to specific time frames, common rediscovery strategies and memories are supported. Users can use their temporal memory as a starting point and then benefit from the session structure and page recognition mechanisms.

4.3.5 Search

Re-searching and keyword searching are among the most common strategies used for re-visitation. These strategies involve either re-creating a previous search or using memories from the targeted page to start a new search. Currently, both re-searching and keyword searching are poorly supported by browsers and users often encounter a series of problems [37]. When re-searching, it is generally difficult to re-create the exact same query as before. As browsers and search engines only provide very basic support, users often need several attempts to re-create a query [39] and need about the same time rediscovering as they needed for their initial search. This process is ineffective, error-prone and often requires significant effort to succeed. Keyword searching is an alternative strategy that is also commonly used for pages that were not accessed by search engines. In this case, users use a search engine to search for keywords they remember from the page they want to re-find. This approach represents a completely new effort, as users start a new search that they have never done before. Depending on the user's memory, searching for keywords is very ineffective and unlikely to succeed. False or fuzzy memories make this process even more difficult, especially because search engines do not take the user's history into account. Alternatively, the address bar can be used to search through the history, but this requires the user to remember parts of the page's title. All in all, both re-searching and keyword searching are time-consuming, error-prone and represent a high mental effort.

Research has shown that a searchable query history can be an effective way to improve re-searching [37]. For this reason, the prototype's search function can be used to search for queries that the user has done before. When the user types, previous queries and visited pages will be suggested below the input field (see Figure 4.16). To support common

Figure 4.16: The search can be used to search through previous queries and pages. Using different words or a different word order still generates useful suggestions

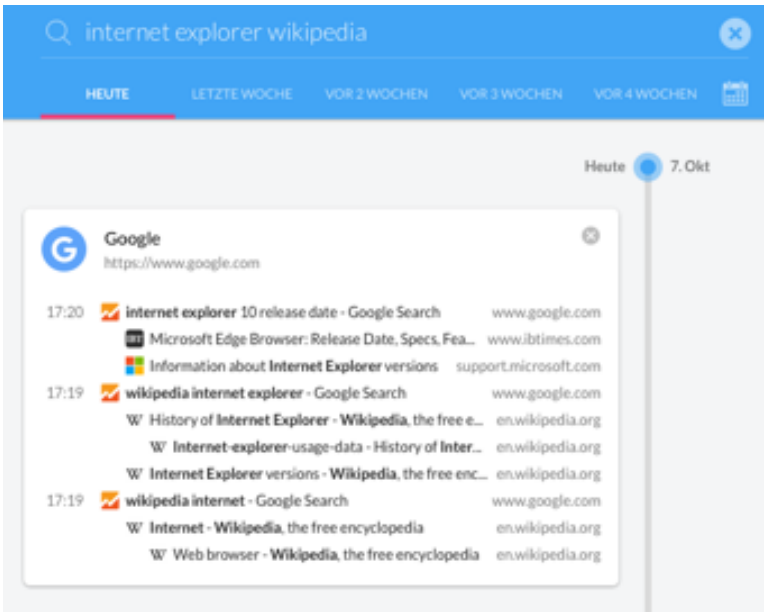

Figure 4.17: When searching, complete sessions are shown as results and words from the query are highlighted.

rediscovery time frames, reduce clutter and prevent irrelevant results, the suggestions do not range back more than 90 days. Moreover, the search takes into account that users are unlikely to enter the exact query they used in the past. Instead of matching previous queries exactly, the best match will be shown depending on the number of matching words. By doing this, it is possible to enter the words of the query in an arbitrary order, while unmatched words will be ignored. Therefore, the correct query will still be suggested if the user enters the query using different words in a different order. This feature directly supports users' approach when re-searching, but does not require the user to try various queries. Even when the query is entered in different ways, the suggestion can be used to directly access the search results.

When re-searching using a search engine, successfully re-creating the query is only the first step towards a successful rediscovery. Afterwards, users still have to compare visited results and re-create browsing paths if the desired page was not accessed directly from the result page. Furthermore, search sessions often consist of several queries and the user might not remember which exact query lead him to the desired page. Re-creating one query from the previous search session might not be enough and the user has to put in further effort. As the developed session view provides an easy way to comprehend browsing paths and also see adjacent searches, it can be used in conjunction with the presented search function to further improve rediscovery. When a user uses the search function or selects a suggestion, the sessions that match the query are shown *completely*. In other words, users are able too see their complete search session when they search for a previous query. By doing this, the user is able to see all searches from his session, all visited results and also the corresponding browsing paths. From this point, the process of identifying the correct page is greatly simplified:

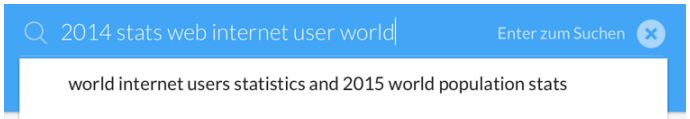

Figure 4.18: Even when the query contains words that are not in the page title or the words are in a different order, there are still useful suggestions. Entering more keywords has no negative side effects and is a valid strategy to improve the suggestions.

1. No additional queries are required, all adjacent searches and results are visible
2. Only pages that were visited are shown
3. Complete browsing paths are visible, including pages that were visited after choosing a search result
4. Matching words from the query are highlighted to quickly identify corresponding parts of the session (see Figure 4.17)
5. The session view provides different ways to recognize the correct page without having to visit it, e.g., thumbnail, path or descriptions

In addition to re-searching, the search feature also provides support when searching for keywords by suggesting the titles of visited pages. After one title is selected, the session that contains this page will be shown. By doing this, the user sees his browsing context and can utilize the page recognition mechanisms. Furthermore, this approach also allows the user to use an arbitrary page from his session as a starting point. Even if he does not remember the title of the page he is looking for, he can search for other pages from his session and locate the desired page afterwards.

When doing a keyword search, there are similar problems as when re-creating queries. Users are unlikely to remember the exact page title and just use arbitrary keywords that they remember about the page. Therefore, the query might contain words that are not in the title or are in a different order. In these cases, the fuzzy search mechanism provides similar advantages as when re-searching. The search suggests the best matching pages while ignoring word order or unmatched words. A common strategy for keyword searching is to refine the search step by step and add additional keywords when there are no useful results. The search feature directly supports this behavior: Entering more keywords improves the suggestions and entering wrong information (i.e., words that are not in the page's title) has no negative side effects (see Figure 4.18).

All in all, the presented search feature directly supports the most common strategies, i.e., re-searching, re-tracing and keyword searching. The suggestions allow the user to only search through queries and page titles from his browsing history, while also supporting inaccurate or fuzzy memories. After searching, users are able to see their complete sessions and context and do not have to do additional queries. Finally, the page recognition mechanisms help to identify pages without having to load them separately.

4.4 Summary

In the beginning of the design process, six objectives for the prototype were determined:

1. Support re-searching
2. Support re-tracing

3. Intelligent keyword search
4. Easy page recognition, e.g., the user should not have to visit all search results again to find the correct one
5. Support for contextual memories
6. Easy to use, little complexity

Re-Searching
The CLIQZ Browsing History provides a search feature which supports users' memories by suggesting previous queries and titles of visited pages below the search bar. Even when the user's query does not match a previous query exactly, the correct query will still be shown as a suggestion. After searching, the whole search session will be shown, including all visited results and related queries. Finally, different search results can be easily distinguished using page previews and other contextual information, such as the browsing path.

This approach directly addresses the most common problems that users encounter while re-searching: query re-creation, page recognition, insufficient memories, path re-creation and changing result pages.

Re-Tracing
Browsers currently provide little support for re-tracing which makes this strategy prone for failure. The CLIQZ Browsing History groups the history into sessions and makes browsing paths visible. Therefore, the user is able to easily comprehend previous sessions. Finding a starting point can be achieved by either using memories of the visit date or by using the search function. This approach allows the user to make use of a range of memories instead of needing to remember information about the target. Searching for an arbitrary page from the session will still result in the complete session. After the user has identified the correct session, he is able to see all visits at a glance. Instead of having to re-create the path step by step and loading each page separately, the target can be accessed directly.

Keyword Search
The reasons for unsuccessful keyword searches are often inaccurate memories and the fact that search engines search the entire web instead of the history. As users have never done such a search before, it represents a completely new effort. To support this strategy, the prototype's search feature also suggests the titles of visited pages. These suggestions do not require the query to match exactly, but rather show the closest match. Furthermore, the suggestions can be improved by adding more keywords to the query. By only showing results from the user's history and supporting inaccurate queries, the effort required for keyword searching is greatly reduced.

Recognition
Besides purple links, current browsers offer no easy way to recognize visited pages. This often becomes a problem when users visited many similar websites, e.g., search results or articles on a shopping site. In this case, the only way to find the correct page is to sequentially try all links until the correct one is recognized. The CLIQZ Browsing History provides a range of features that make page recognition easier. By showing sessions and browsing paths, the history becomes more comprehensible and scannable. Furthermore, there is an extensive preview of pages on mouseover that includes a thumbnail or prominent image and a description of the page. This makes it possible to compare and identify pages without having to load them separately. Instead of only relying on titles,

users are able to make use of more common memories, i.e., visual elements and context.

Contextual Memories
As contextual memories are very common and often remembered better than actual information about the target, I tried to incorporate support for such memories in every feature. Grouping the history into sessions allows to show consecutive visits, while also making browsing paths visible. When searching, the user will always see whole sessions as results. It is possible to use any page from a session as a starting point when searching for a page inside this session. Users are also able to use temporal information as a starting point. Furthermore, by providing a scrollable and scannable interface, sessions can be comprehended at a glance and might trigger further memories about the desired page. All in all, users are able to utilize a range of memories that are not utilizable when rediscovering in current browsers.

Usability
The prototype uses a very simple interface and requires little interaction. The complete history can be accessed by just scrolling up and down. There are few interface elements which are widely known and require no learning. When designing the CLIQZ Browsing History, the goal was to develop a tool that supports current strategies instead of providing alternatives. Users do not have to change their habits as they can still utilize the same strategies, e.g., re-searching or re-tracing. I conducted four usability tests with the final tool which did not show any problems in understanding or using the prototype. The participants were able to grasp the concept of the interface easily and used all features without any problems. To further investigate the performance of the tool, I conducted an extensive evaluation which is explained in the next chapter.

4.5 Implementation

The *CLIQZ Browsing History* was developed as a part of the CLIQZ navigation extension for Firefox. Firefox extensions are written in JavaScript, but are also able to utilize more advanced features, e.g., accessing local files or creating databases.

To develop the prototype itself, I used AngularJS [17], Angular Material [16], jQuery [46], jQuery UI [47] and other smaller JavaScript libraries. Angular utilizes the Model-view-controller (MVC) [53] pattern.

4.5.1 Model

The history that is provided by Firefox does not contain enough information to re-create browsing paths. Therefore, I developed an advanced history database that is created when using the CLIQZ extension. This SQL database contains the user's complete history and additionally saves data that is needed to reconstruct sessions. For example, this history database also stores queries, when and how sessions are started (bookmark, search, history, etc.), how the user interacts with websites, how much time is spent on a page and lastly, how individual page visits are linked so it's possible to re-create the browsing path.

Furthermore, OpenGraph data and thumbnails of pages are automatically generated and saved while a page is visited. By doing this, the thumbnails and preview information can be displayed immediately in the history view.

Before the sessions are displayed, they are processed and filtered. It might be necessary to merge several branches of the browsing tree when they represent the same page. Furthermore, some pages are automatically removed from the session when there was no interaction and the user spent very little time on the page, e.g., redirects.

4.5.2 View and Controller

The main interface of the CLIQZ Browsing History is implemented as a normal website using HTML and CSS. There are three Angular controllers which control different parts of the interface. The first controller handles the history view and displays the different browsing sessions. As reconstructing sessions from the history is time-consuming, the processed sessions are completely cached in the SQL history database. The logos and favicons are all loaded from a single CLIQZ server that has already been used in the extension before. When the page is scrolled to the bottom, new sessions are dynamically loaded from the database.

The second controller is responsible for showing the previews on mouseover. As the OpenGraph data and thumbnails are already saved in the user's profile directory, they do not have to be loaded separately and can be displayed immediately. Only images that are linked in the OpenGraph data will be loaded on demand.

Finally, the third controller handles the header, i.e., the search function and clicks on the buttons for temporal navigation. When one of these buttons is clicked, the history will scroll automatically and load all sessions up to that point in time dynamically. When a user enters a query into the search bar, the suggestions are generated on demand using the SQL history database. When a search is performed, only sessions that match all of the words from the query will be shown. The history view is automatically scrolled to the first matching session. Finally, search queries are also stored in the URL. Therefore, the query will be repeated automatically when a user clicks on a page and uses the back button afterwards.

4.5.3 Summary

The *CLIQZ Browsing History* relies on an advanced history database that contains additional data to reconstruct browsing sessions. However, due to a high modularization, it is easily possible to port the CLIQZ Browsing History to different browsers in the future. To do this, only the history interface that is provided by the browser has to be adjusted.

All in all, I used modern technologies to create a future-proof system that can be easily extended in the future.

5 Evaluation of the CLIQZ Browsing History

To find out how the *CLIQZ Browsing History* performs, I conducted a study where it was used to rediscover websites and web pages. The goals of this study include (1) assessing the usability, (2) analyzing the user interaction and (3) comparing the performance to normally used revisitation strategies.

5.1 Hypotheses

- H_1: Participants using the CLIQZ Browsing History need less time to rediscover pages
- H_2: Participants using the CLIQZ Browsing History have a higher success rate
- H_3: Participants using the CLIQZ Browsing History need less page visits to rediscover pages
- H_4: Participants using the CLIQZ Browsing History mostly use the search feature as it supports the most common strategies

5.2 Dependent and Independent Measures

The study followed a between subjects design with one independent variable, the method used to rediscover pages:

1. The CLIQZ Browsing History
2. The user's normal browsing environment with tools and strategies of their choice

The dependent variables include time needed for rediscovery, number of page visits and success rate. Furthermore, a questionnaire is used to determine users' problems and strategies and the usability of the CLIQZ Browsing History is assessed using the System Usability Scale.

5.3 Study Design

The study is a 2 x 10 (conditions x repetitions) between-subjects design (2 x 5 for external participants) and consists of three parts. Participants who were not associated with CLIQZ had to less tasks to reduce the needed time and increase the willingness to participate. It is run unsupervised on the participants own computers, using their normal browsing environment and history. The first part is a questionnaire that asks about general information, computer skill level and also how web pages are rediscovered, i.e., what strategies are used and what problems are encountered:

1. Age (optional)
2. Hours of Internet usage per day
3. Computer skill level
4. Revisitation frequency
5. Revisitation success
6. Problems (5 point Likert scale)
   (a) Remembering query
   (b) Remembering enough information about the page to initiate a search

Figure 5.1: The main part of the study shows participants a blurred screenshot of a page from their browsing history and a textual description of the content. Users have to re-find this page using either the CLIQZ Browsing History or a strategy of their choice.

   (c) Changing page contents
   (d) Rediscovery taking a long time
   (e) Other (free text)
7. Strategies (explained with examples, 5 point Likert scale)
   (a) Re-searching
   (b) Keyword searching
   (c) Browser history
   (d) Using address bar to search through history
   (e) Re-tracing
   (f) Other (free text)

In the second part, participants are presented with pages from their browsing history that they have to re-visit using either the CLIQZ Browsing History (first condition) or a strategy of their choice (second condition). These tasks are generated by an algorithm that selects visited pages from the last three weeks that were only visited once and are likely to be revisited. These pages are chosen by a range of metrics, such as type (article, search result, video), time spent and session length. Participants are first shown a blurred screenshot of the page and if available, a description from the Opengraph data or, alternatively, keywords generated from the page title (see Figure 5.1). This information represents common memories of visited pages, i.e., visual structure and keywords from the content. If the user remembers visiting the page, he is presented with the task to re-find it. For this purpose, a new tab with either (1) the CLIQZ Browsing History or (2) an empty page is opened and the user is asked to return to the study when he finds the page or feels like he is not able to find it. To complete the second part, participants have to do ten (five for external participants) rediscoveries, independent of their success. If the user chooses to skip a task because the page is private or he does not remember visiting it, it does not count towards the progress.

After completing the revisitation tasks, the participants have to fill out a final questionnaire where they can give general feedback about the study and, if applicable, are able to rate the CLIQZ Browsing History using the System Usability Scale (SUS) [48].

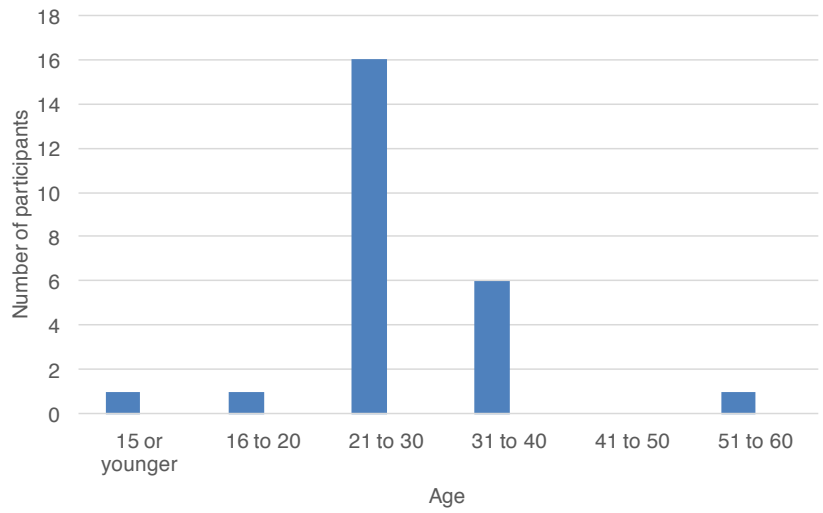

Figure 5.2: Participants' age distribution

5.4 Participants

Participants for this study were randomly selected users of the CLIQZ Firefox extension. While browsing, a message appeared that asked them whether they wanted to participate. This message was only shown to users that had enough browsing history to generate tasks for the study. Furthermore, employees of CLIQZ were also able to participate. All participants entered into a raffle where they were able to win an Amazon gift card.

Ultimately, 28 participants were selected for this study, consisting of 9 external CLIQZ users and 19 CLIQZ employees. 103 data sets were collected for the first (14 participants) and 66 for the second condition (14 participants). Users that dropped out in the second part, but still completed at least one task are also included in the results. On average, participants in the first condition (CLIQZ Browsing History) completed 7.4 (max 10) tasks and participants in the second condition (user strategy) completed 4.7 (max 10) tasks. Two data sets were removed because participants in the first condition did not use the CLIQZ Browsing History.

5.5 Data Collection

The CLIQZ extension has full access to the user's browsing history. During the second part, the following data was collected:

- All websites and pages visited during the revisitation tasks
- Queries and address bar inputs
- Time spent on visited pages
- Interaction with pages, i.e., number of clicks, keystrokes and scrolling
- Start and end time of task

Additionally, when using the history tool, the following usage data was measured:

- Number of clicks on header links (week selection) and calendar
- Queries and chosen suggestions
- Number of viewed days from history
- Pages that are visited
- Number of shown link previews
- Time spent in history tool

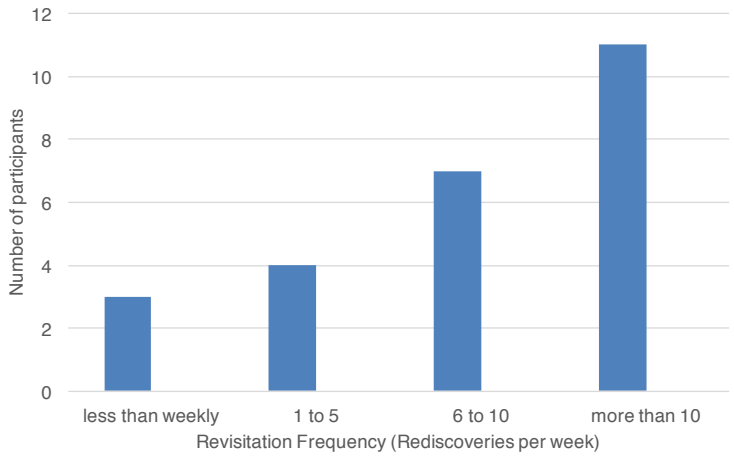

Figure 5.3: Users rediscover pages from their history frequently, most users perform more than 10 rediscoveries per week.

5.6 Results

5.6.1 Revisitation Behavior

Most of the participants who entered their age are in the range of 21 to 30, followed by the group of 31 to 40 year olds (see Figure 5.2). The average Internet usage is about 7.3 hours per day. Only one participant stated a computer skill level below average, while 32% stated an average and 67% an above average skill level.

When asked for the frequency of rediscoveries, most participants reported rediscovering pages more than ten times per week (46%), while 28% do less than five rediscoveries per week (see Figure 5.3).

Rediscoveries are mostly successful: 72% stated that they are often or very often able to rediscover pages, while 28% stated that they fail at least sometimes. The most common problems while rediscovering pages are remembering the query and remembering enough information about the page to initiate a search. Two participants explicitly mentioned that they are often unable to remember the title of a page and have no effective way of utilizing other memories.

Re-searching and keyword searching are the most commonly used strategies, utilized often or very often by 72% and 76% of participants respectively. The browser's history is the least used strategy, 64% stated that they use it only rarely or very rarely.

These findings coincide with related work and the results previously reported in this work. Users frequently rediscover content from their history and although they are mostly successful, they still encounter a range of problems. Instead of using browser tools, re-searching and keyword searching are the most common strategies. When doing this, users often have problems remembering their queries or initiating a search and cannot make use of contextual memories.

5.6.2 Revisitation Performance

Time

To assess the performance of the developed tool, I compared the time that users needed to successfully rediscover pages. When using the CLIQZ Browsing History, the average revisitation time was 23.5 seconds (SEM = 3.3), while users who used a strategy of their choice needed 26.9 seconds (SEM = 4.2) (see Figure 5.4). An unpaired t-test did not show any significance ($t_{13} = 0.63$, $p = 0.53$), so H_1 is not supported. Therefore, it can be assumed that both conditions perform similarly for the time that is needed to rediscover a page from the history.

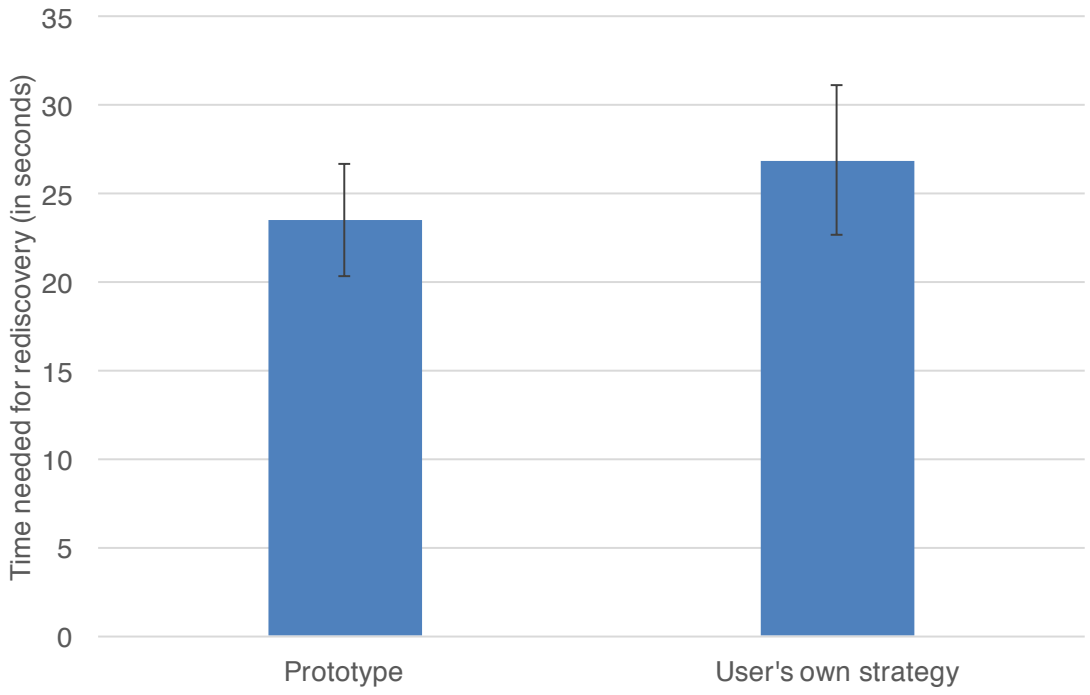

Figure 5.4: Time needed for rediscovery (mean ± SEM). Prototype n = 14 participants; user's strategy n = 14 participants

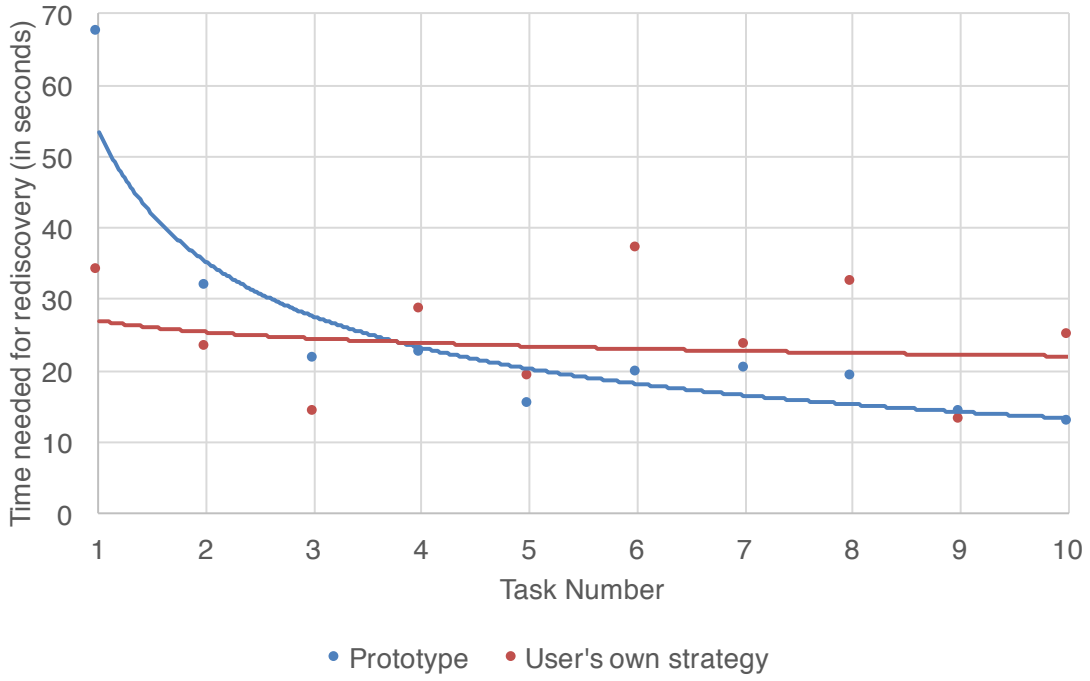

Figure 5.5: Users who used the CLIQZ Browsing History for rediscovery performed poor in the first two tasks and performed better for later tasks. This learning effect cannot be observed for the second condition as users used familiar strategies.

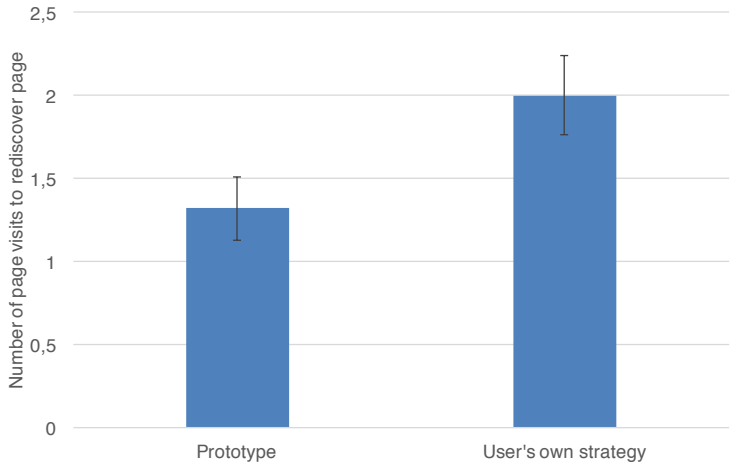

Figure 5.6: Page visits needed for rediscovery (mean ± SEM). Prototype n = 14 participants; user's strategy n = 14 participants

As users use CLIQZ Browsing History for the first time, it can be assumed that there is a learning effect which causes the first tasks to take longer than the following ones. To investigate this, I looked at the average rediscovery times per task (see Figure 5.5). Users who used the CLIQZ Browsing History tended to need less time over the course of the study and the first and second task performed comparably poor. After doing two rediscoveries using the CLIQZ Browsing History, the performance leveled out and remained constant. On the other hand, the second condition performed about equal for all tasks. This is not surprising as users are able to use strategies that they would normally use and therefore require no learning. Although these findings show a promising performance increase, it is difficult to infer any further conclusions. The second condition had less tasks per user (4.7 compared to 7.4) and therefore less data for later tasks. Therefore, it is not possible to make any statements about the performance after users became familiar with the history tool.

All in all, the performance of the *CLIQZ Browsing History* shows the promise of the underlying concept. Even when not accounting for the learning effect, users performed slightly better in terms of rediscovery time. Future studies can focus on long-term users of the prototype who already became familiar with its functionality. Furthermore, the history tool still has performance issues when searching which might have slowed users down. Further improvements can therefore increase the performance and make the prototype more effective.

Success Rate

Another performance metric is the success rate of rediscoveries. Users who used the CLIQZ Browsing History failed in 11.8% of rediscoveries, while users who used their own strategies failed in 14.3% of rediscoveries. This difference is not significant ($\chi^2 = 0.09$, $p = 0.77$) and H_2 is unsupported by the results. However, there is still room for improvement in the prototype. In at least one case there was a bug that did not show the user's complete history and therefore prevented a successful rediscovery. Furthermore, users might need some time to get used to some features of the prototype. As the search feature works differently than browsers' history searches, users potentially need some time to get used to the CLIQZ Browsing History, how it works and what memories they can use to start a search. At this point, there is not enough data to make definite statements about the prototype's success rate, but further improvements and more extensive studies could provide further insights.

Page Visits needed for Rediscovery
Another indicator for the performance of a rediscovery is the number of page visits that is needed to revisit a page. Common methods such as re-searching and re-tracing involve re-creating previous browsing paths and therefore require a high number of page visits. However, visiting a higher number of pages increases the needed time and mental effort, as each step requires scanning and analyzing the current page. The prototype was designed to minimize page visits and provide the user with direct links to every page from his history.

When users used the CLIQZ Browsing History, they needed 1.32 page visits (SEM = 0.19) to successfully rediscover a page, while users needed 2.00 page visits (SEM = 0.24) when using their own strategies. An unpaired t-test shows a significant difference (t_{13} = 2.2, p = 0.037) and therefore, H_3 is accepted. This shows that the CLIQZ Browsing History provides shortcuts and users rely less on re-tracing. Furthermore, the prototype's recognition features prove to be effective, as users rarely visit a wrong page and then go back to the prototype. The observed value of 1.32 is close to a "perfect" value of 1.00 which would mean that users always rediscover a page by direct access.

All in all, the prototype allows users to identify pages without loading them and users are able to rediscover pages directly. Users were provided with enough information and confidence that they did not have to rely on re-tracing pages. Therefore, the prototype reduces the mental effort and also provides a single interface for all strategies.

5.6.3 Usability

Participants who used the developed prototype rated it using the System Usability Scale (SUS). Overall, the *CLIQZ Browsing History* scores 75 points which is considered above average [48].

5.6.4 User Behavior

During the study, a range of usage data was collected, including interaction with pages (clicks, keystrokes, scrolls), all page visits during the tasks and interaction with the prototype. Most of this data was used as debugging information and was not analyzed in detail. However, a future analysis could provide further insights on user strategies and the interaction with the CLIQZ Browsing History.

To understand how participants interacted with the CLIQZ Browsing History, I looked at how often every feature was used by each user (see Figure 5.7). H_4 is accepted, as most participants used the search feature almost exclusively (82% of all tasks), mostly in combination with suggestions. The navigational links in the header were barely used (4% of all tasks), but users regularly scrolled through the history manually. Previews were only utilized in 18% of all tasks, which shows that the session representation is mostly sufficient for recognizing the pages. One user only needed 4.3 seconds on average, but still used features of the prototype. I assume that this user stated that his rediscovery was successful after only briefly interacting with the prototype and not actually finding the page.

This data supports the initial assumption that search is the most used feature because it supports important rediscovery strategies, such as re-searching and keyword searching. As these strategies were also rated as the most common by participants (used often or very often by 72% and 76% of users), these results show that users were still able to utilize these familiar strategies. Furthermore, suggestions were used in most tasks (55% of all tasks), which shows that they provide further support in rediscovering pages. While some users relied on scrolling, the navigational links were mostly unused. One user stated that his

Searches	Suggestions	Scrolls	Navigation	Previews	Time
100%	100%	13%	0%	13%	38.5
100%	80%	40%	0%	30%	13.6
0%	0%	20%	20%	20%	4.3
78%	22%	44%	0%	11%	24.8
0%	0%	75%	0%	0%	22.5
71%	71%	0%	14%	0%	20.3
100%	78%	22%	0%	33%	33.7
100%	44%	33%	11%	22%	35.9
100%	43%	57%	0%	0%	15.2
100%	50%	50%	0%	50%	49.9
100%	71%	29%	0%	29%	21.9
100%	80%	30%	0%	20%	20.8
100%	71%	0%	14%	14%	17.8
100%	57%	0%	0%	14%	10.5
82%	55%	29%	4%	18%	22.4

Figure 5.7: Each row represents a participant that used the prototype and shows how often each feature was used (tasks in which feature was used divided by all tasks done by user).

memory of time is often inaccurate and not sufficient to use as a starting point. Furthermore, users probably are not accustomed to using temporal information for rediscoveries, as it is poorly supported by browsers. Lastly, only 18% of tasks involved the usage of previews. As users were still able to successfully rediscover pages and only needed 1.32 page visits on average, it can be assumed that the session representation already provides enough information in most cases.

5.7 Summary

Although the presented data shows no significant differences for rediscovery time and success rate, the evaluation still supports the underlying concepts. Furthermore, I observed a promising performance increase in later trials, so a future study of long-term users could show further benefits.

There was a significant difference for the number of page visits that users needed to rediscover pages. This shows that the history presentation and recognition features are effective, as users need to visit less pages to identify the correct target. Instead, they are able to use the prototype interface to visit pages directly.

When looking at the interaction with the CLIQZ Browsing History, I found that 82% of tasks involved the usage of the search feature. Participants were able to use familiar strategies, such as re-searching and keyword searching. In 67% of tasks that involved the search feature, users also used the search suggestions. Therefore, participants were directly supported in their intends and had to rely less on their memories. By providing these suggestions and supporting all strategies, users are able to use a single interface and the mental effort is reduced.

Finally, users were able to quickly learn how to use the CLIQZ Browsing History and it achieved an above average score on the System Usability Scale (SUS).

5.8 Limitations

As CLIQZ is currently only available for Firefox, only Firefox users participated in the study. Although all browsers offer similar revisitation tools, there are slight differences which might cause different behaviors or strategies.

A further limitation of the study is caused by the automatic page selection from the users' history and the automated and unsupervised approach. It is difficult to predict pages that are worth rediscovering and represent them in a way that is similar to users' memories. One user reported that the blurred screenshots of pages helped him to gain memories that he would not have had otherwise. As the average revisitation times were below one minute, it can be assumed that the chosen pages were relatively easy to rediscover. In reality, rediscoveries are often a very complex and more time-consuming task. Furthermore, the circumstances in which the study was performed differed across users.

Although I observed a promising performance increase in later trials, it was not possible to draw any conclusions because of the limited amount of data. There was an imbalance in tasks per user (7.4 for first and 4.7 for second condition) which made it difficult to make definite conclusions.

As thumbnails were only available to users that used a special version of the CLIQZ extension, some users had to rely on the OpenGraph data for recognition. Finally, there were some bugs and performance issues in the history prototype which might have caused longer revisitation times and more failures.

6 Conclusion and Future Work

In this work, I presented *CLIQZ Browsing History*, a history interface that improves website revisitation by supporting common user strategies, behaviors and memories. The evaluation proved the underlying concepts and showed promising performance increases after continued usage. The developed prototype was easily usable and required little mental effort, while significantly lowering the number of page visits needed for successful rediscoveries.

6.1 Learnings and Recommendations

The data from my initial studies and evaluation showed that revisitation and rediscovery are still frequent tasks that are often difficult and time-consuming. Users still encounter the same problems that were explored in past studies. Current browsers offer poor support for revisitation and users rely on re-searching, re-tracing and keyword searching. These strategies require high mental effort as users need to have exact memories of the query, page title or URL. The fact that users need about the same time to rediscover a page as they initially needed to find it, shows that revisitation still has great potential for improvement.

As browsers do not have a specific target audience, it was important to create an interface that is easy to use. I achieved this by reducing interaction, using known concepts (search, tree view, timeline) and also minimizing the amount of information on the screen by showing more details on mouseover. In the first design approaches of the prototype I created an interface where it was possible to interactively navigate previous browsing sessions. In the end, I settled with a simpler interface that requires as little interaction as possible. This design choice was confirmed in usability tests and in the evaluation. While the interactive approach was often difficult to grasp, the final prototype reached a high score on the System Usability Scale and the evaluation showed that participants got used to the prototype after two to three trials. Also, users of the prototype relied almost exclusively on search as it supports familiar strategies. In summary, when designing interfaces for everyone, I recommend (1) reducing interaction, (2) using known concepts, (3) focusing on only the most important features and (4) keeping the main interface as simple as possible, e.g., by showing information on demand.

To maximize the usability of the *CLIQZ Browsing History*, I tried to provide support for strategies that users are already familiar with, e.g., re-searching and keyword searching. Furthermore, I implemented features to support memories that are not usable with current browser tools. Using the CLIQZ Browsing History, it is possible to use temporal or contextual information as a starting point for a search. However, I found that users almost exclusively use strategies that they are used to. As a consequence, I recommend to always analyze user behavior comprehensively, especially when designing alternatives to familiar behaviors. Convincing users to use alternatives can be very difficult and can therefore require additional measures when designing interfaces.

During user studies, I found that users often prefer a longer, but known, browsing path over a shortcut because they have more confidence when following a known path. Therefore, it is not only important to show the correct search results, but also convince the user that a specific result represents the page he is looking for. To achieve this in my prototype, search results always consist of complete sessions instead of single pages. As a result, users who used the CLIQZ Browsing History needed a significantly shorter browsing path to reach their targets. Showing the context of search results not only helped

users to directly access their targets, but also reduced the required mental effort as it was not necessary to look at many separate pages. By providing a single, scannable interface and showing context, it was possible to convince users to use shortcuts. Therefore, when designing a search interface, I recommend to not only focus on showing the best results, but also analyze how contextual information can give users a higher confidence.

6.2 Future Work

The evaluation of the *CLIQZ Browsing History* raised some questions and revealed further aspects of the prototype that are worth exploring.

In my evaluation, the average revisitation time was under one minute, i.e., the tasks were often not difficult for users. In the future, a study with more complex rediscoveries that better represent common scenarios could give further insights. As it is difficult to automatically generate rediscovery tasks, it is worth looking at how to represent pages and how to create more realistic tasks.

When using the prototype, users needed significantly less page visits to rediscover content. Therefore, it can be assumed that the CLIQZ Browsing History improves recognition and offers effective shortcuts. Users rarely clicked on a wrong link and were therefore able to identify correct pages with the displayed information. This can mostly be attributed to the session view and visibility of browsing paths, as only few users utilized the mouseover previews. To further improve page recognition, it can be worth looking at how the previews can be made more prominent and be enhanced to better represent common memories of web pages.

I collected a range of usage data for all users of the prototype, which was mostly used as debugging information. Future analysis of this data could give better insights on how people interacted with the prototype, how they used specific features and what strategies they used. By doing this, the features of the CLIQZ Browsing History could be improved and be better tailored to users' strategies.

Users almost exclusively used the search feature when rediscovering pages with the CLIQZ Browsing History. On the one hand, users may not be familiar with using contextual memories as starting points for their search or, on the other hand, they might prefer to stick to familiar strategies. It is worth investigating how certain features can be promoted or how users can be taught different revisitation strategies that can be more effective.

Although there was no significant difference in the revisitation time and success rate, I observed a performance increase in later tasks for participants that used the prototype. Because of a lack of data, I was not able to make any conclusions. It is worth exploring how users perform when they are already used to the features of the prototype and use it regularly. Furthermore, there are still bugs and performance issues that offer room for improvement. As the prototype already performed similar to strategies that users are familiar with, there might be a significant performance increase for long-term users.

Inhalt der beigelegten CD

- Arbeit als PDF, sowie LaTeX-Dateien
- Antritts- und Abschlussvortrag als PDF
- Related Work
- Ergebnisse der Studien und der Evaluation als JSON
- Python-Skripte, die zur Auswertung genutzt wurden
- Quellcode der CLIQZ Browsing History, sowie der durchgeführten Evaluation
- Quellcode der gesamten CLIQZ Firefox Erweiterung (inklusive der CLIQZ Browsing History)
- Installierbares Firefox-Addon (inklusive der CLIQZ Browsing History)

References

[1] D. Abrams and R. Baecker. How people use www bookmarks. In *CHI '97 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '97, pages 341–342, New York, NY, USA, 1997. ACM.

[2] E. Adar, J. Teevan, and S. T. Dumais. Large scale analysis of web revisitation patterns. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '08, pages 1197–1206, New York, NY, USA, 2008. ACM.

[3] E. Agichtein, R. W. White, S. T. Dumais, and P. N. Bennet. Search, interrupted: Understanding and predicting search task continuation. In *Proceedings of the 35th International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '12, pages 315–324, New York, NY, USA, 2012. ACM.

[4] A. Aula, N. Jhaveri, and M. Käki. Information search and re-access strategies of experienced web users. In *Proceedings of the 14th International Conference on World Wide Web*, WWW '05, pages 583–592, New York, NY, USA, 2005. ACM.

[5] A. Aula, R. M. Khan, Z. Guan, P. Fontes, and P. Hong. A comparison of visual and textual page previews in judging the helpfulness of web pages. In *Proceedings of the 19th International Conference on World Wide Web*, WWW '10, pages 51–60, New York, NY, USA, 2010. ACM.

[6] E. Barrett, editor. *Text, Context, and Hypertext: Writing with and for the Computer*. MIT Press, Cambridge, MA, USA, 1988.

[7] T. Blanc-Brude and D. L. Scapin. What do people recall about their documents?: Implications for desktop search tools. In *Proceedings of the 12th International Conference on Intelligent User Interfaces*, IUI '07, pages 102–111, New York, NY, USA, 2007. ACM.

[8] R. L. Brown, N.R. and S. Shevell. The subjective dates of natural events in very-long-term memory. *Cognitive Psychology*, 1985.

[9] J. W. Bruce, H. and S. Dumais. Keeping and re- finding information on the web: What do people do and what do they need? *Proceedings of ASIST '04*, 2004.

[10] D. B. Choo, C.W. and D. Turnbull. Information seeking on the web: An integrated model of browsing and searching. *American Society for Information Science*, 2000.

[11] A. Cockburn and S. Greenberg. Issues of page representation and organisation in web browser's revisitation tools. *Australasian Journal of Information Systems, Vol. 7, No 2*, 2000.

[12] A. Cockburn and B. McKenzie. What do web users do? an empirical analysis of web use. *International Journal of Human-Computer Studies*, 2001.

[13] S. Dumais, E. Cutrell, J. Cadiz, G. Jancke, R. Sarin, and D. C. Robbins. Stuff i've seen: A system for personal information retrieval and re-use. In *Proceedings of the 26th Annual International ACM SIGIR Conference on Research and Development in Informaion Retrieval*, SIGIR '03, pages 72–79, New York, NY, USA, 2003. ACM.

[14] S. Dziadosz and R. Chandrasekar. Do thumbnail previews help users make better relevance decisions about web search results? In *Proceedings of the 25th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '02, pages 365–366, New York, NY, USA, 2002. ACM.

[15] Facebook. The Open Graph protocol. http://www.ogp.me/, 10 2014.

[16] Google. Angular Material. https://material.angularjs.org/latest/, 10 2015.

[17] Google. AngularJS. https://angularjs.org/, 10 2015.

[18] J. Hailpern, N. Jitkoff, A. Warr, K. Karahalios, R. Sesek, and N. Shkrob. Youpivot: Improving recall with contextual search. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '11, pages 1521–1530, New York, NY, USA, 2011. ACM.

[19] International Telecommunication Union (ITU). Individuals using the internet 2005 to 2014. http://www.itu.int/en/ITU-D/Statistics/Documents/statistics/2014/ITU_Key_2005-2014_ICT_data.xls, 05 2015.

[20] B. Jiao, L. Yang, J. Xu, and F. Wu. Visual summarization of web pages. In *Proceedings of the 33rd International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '10, pages 499–506, New York, NY, USA, 2010. ACM.

[21] W. Jones. Once found, what then? a study of "keeping" behaviors in the personal use of web information. *Proc. of the American Society for Information Science and Technology*, 2002.

[22] S. Kaasten and S. Greenberg. Integrating back, history and bookmarks in web browsers. In *CHI '01 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '01, pages 379–380, New York, NY, USA, 2001. ACM.

[23] R. Kawase, G. Papadakis, E. Herder, and W. Nejdl. Beyond the usual suspects: Context-aware revisitation support. In *Proceedings of the 22Nd ACM Conference on Hypertext and Hypermedia*, HT '11, pages 27–36, New York, NY, USA, 2011. ACM.

[24] C. Kuhlthau. Inside the search process: Information seeking from the user's perspective. *Journal of the American Society for Information Science*, 1991.

[25] J.-S. Lee, D. Tatar, and E. R. Pedersen. Time, topic and trawl: Stories about how we reach our past. In *Proceedings of the Designing Interactive Systems Conference*, DIS '12, pages 234–243, New York, NY, USA, 2012. ACM.

[26] S. LeeTiernan, S. Farnham, and L. Cheng. Two methods for auto-organizing personal web history. In *CHI '03 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '03, pages 814–815, New York, NY, USA, 2003. ACM.

[27] L. A. Leiva, V. J. Traver, and V. Castelló. Livethumbs: A visual aid for web page revisitation. In *CHI '13 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '13, pages 1797–1802, New York, NY, USA, 2013. ACM.

[28] I. Li, J. Nichols, T. Lau, C. Drews, and A. Cypher. Here's what i did: Sharing and reusing web activity with actionshot. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '10, pages 723–732, New York, NY, USA, 2010. ACM.

[29] J. Liu, P. J. Hong, and E. R. Pedersen. Research trails: Getting back where you left off. In *Proceedings of the 19th International Conference on World Wide Web*, WWW '10, pages 1151–1152, New York, NY, USA, 2010. ACM.

[30] B. Ma Kay and C. Watters. Exploring multi-session web tasks. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '08, pages 1187–1196, New York, NY, USA, 2008. ACM.

[31] B. MacKay, M. Kellar, and C. Watters. An evaluation of landmarks for re-finding information on the web. In *CHI '05 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '05, pages 1609–1612, New York, NY, USA, 2005. ACM.

[32] B. MacKay and C. Watters. Building support for multi-session tasks. In *CHI '09 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '09, pages 4273–4278, New York, NY, USA, 2009. ACM.

[33] G. R. Mary P. Czerwinski, Maarten van Dantzich and H. Hoffman. The contribution of thumbnail image, mouse-over text and spatial location memory to web page retrieval in 3d. *INTERACT '99*, 1999.

[34] P. Mathur and K. Karahalios. Using bookmark visualizations for self-reflection and navigation. In *CHI '09 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '09, pages 4657–4662, New York, NY, USA, 2009. ACM.

[35] N. Matthijs and F. Radlinski. Personalizing web search using long term browsing history. In *Proceedings of the Fourth ACM International Conference on Web Search and Data Mining*, WSDM '11, pages 25–34, New York, NY, USA, 2011. ACM.

[36] D. Morris, M. Ringel Morris, and G. Venolia. Searchbar: A search-centric web history for task resumption and information re-finding. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '08, pages 1207–1216, New York, NY, USA, 2008. ACM.

[37] H. Obendorf, H. Weinreich, E. Herder, and M. Mayer. Web page revisitation revisited: Implications of a long-term click-stream study of browser usage. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '07, pages 597–606, New York, NY, USA, 2007. ACM.

[38] V. L. O'Day and R. Jeffries. Orienteering in an information landscape: How information seekers get from here to there. In *Proceedings of the INTERACT '93 and CHI '93 Conference on Human Factors in Computing Systems*, CHI '93, pages 438–445, New York, NY, USA, 1993. ACM.

[39] H.-T. Pu and X.-Y. Jiang. A comparison of how users search on web finding and re-finding tasks. In *Proceedings of the 2011 iConference*, iConference '11, pages 446–451, New York, NY, USA, 2011. ACM.

[40] S. G. Shaun Kaasten and C. Edwards. How people recognize previously seen web pages from titles, urls and thumbnails. *Proceedings of HCI '02*, 2002.

[41] C. Staff and I. Bugeja. Automatic classification of web pages into bookmark categories. In *Proceedings of the 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, SIGIR '07, pages 731–732, New York, NY, USA, 2007. ACM.

[42] L. Tauscher and S. Greenberg. How people revisit web pages: empirical findings and implications for the design of history systems. *Int. J. Hum.-Comput. Stud.*, 1997.

[43] L. Tauscher and S. Greenberg. Revisitation patterns in world wide web navigation. In *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems*, CHI '97, pages 399–406, New York, NY, USA, 1997. ACM.

[44] J. Teevan. How people recall, recognize, and reuse search results. *ACM Trans. Inf. Syst.*, 26(4):19:1–19:27, Oct. 2008.

[45] J. Teevan, E. Cutrell, D. Fisher, S. M. Drucker, G. Ramos, P. André, and C. Hu. Visual snippets: Summarizing web pages for search and revisitation. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '09, pages 2023–2032, New York, NY, USA, 2009. ACM.

[46] The jQuery Foundation. jQuery. `https://jquery.com/`, 10 2015.

[47] The jQuery Foundation. jQuery UI. `https://jqueryui.com/`, 10 2015.

[48] U. S. Department of Health Human Services. System Usability Scale (SUS). `http://www.usability.gov/how-to-and-tools/methods/system-usability-scale.html`, 10 2015.

[49] Q. Wang and H. Chang. Multitasking bar: Prototype and evaluation of introducing the task concept into a browser. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '10, pages 103–112, New York, NY, USA, 2010. ACM.

[50] H. Weinreich, H. Obendorf, E. Herder, and M. Mayer. Off the beaten tracks: Exploring three aspects of web navigation. In *Proceedings of the 15th International Conference on World Wide Web*, WWW '06, pages 133–142, New York, NY, USA, 2006. ACM.

[51] H. Weinreich, H. Obendorf, E. Herder, and M. Mayer. Not quite the average: An empirical study of web use. *ACM Trans. Web*, 2(1):5:1–5:31, Mar. 2008.

[52] J. Wen. Post-valued recall web pages: User disorientation hits the big time. *IT Society*, 2003.

[53] Wikipedia. Model–view–controller. `https://en.wikipedia.org/wiki/Model%E2%80%93view%E2%80%93controller`, 10 2015.

[54] S. S. Won, J. Jin, and J. I. Hong. Contextual web history: Using visual and contextual cues to improve web browser history. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '09, pages 1457–1466, New York, NY, USA, 2009. ACM.

[55] C. Ye, M. L. Wilson, and T. Rodden. Develop, implement, and improve a web session detection model. In *Proceedings of the 5th Information Interaction in Context Symposium*, IIiX '14, pages 336–338, New York, NY, USA, 2014. ACM.