"WEBSCRAPS" – A TOOL TO MANAGE
WEB INFORMATION GATHERING TASKS

by

Timothy Mankowski

Submitted in partial fulfillment of the requirements
for the degree of Master of Computer Science

at

Dalhousie University
Halifax, Nova Scotia
April 2011

DALHOUSIE UNIVERSITY

FACULTY OF COMPUTER SCIENCE

The undersigned hereby certify that they have read and recommend to the Faculty of Graduate Studies for acceptance a thesis entitled ""WEBSCRAPS" – A TOOL TO MANAGE WEB INFORMATION GATHERING TASKS" by Timothy Mankowski in partial fulfillment of the requirements for the degree of Master of Computer Science.

Dated: April 13, 2011

Supervisor: _____

Readers: _____

_____

DALHOUSIE UNIVERSITY

DATE: April 13, 2011

AUTHOR: Timothy Mankowski

TITLE: "WEBSCRAPS" – A TOOL TO MANAGE WEB INFORMATION
GATHERING TASKS

DEPARTMENT OR SCHOOL: Faculty of Computer Science

DEGREE: M.C.Sc. CONVOCATION: October YEAR: 2011

Permission is herewith granted to Dalhousie University to circulate and to have copied for non-commercial purposes, at its discretion, the above title upon the request of individuals or institutions. I understand that my thesis will be electronically available to the public.

The author reserves other publication rights, and neither the thesis nor extensive extracts from it may be printed or otherwise reproduced without the author's written permission.

The author attests that permission has been obtained for the use of any copyrighted material appearing in the thesis (other than the brief excerpts requiring only proper acknowledgement in scholarly writing), and that all such use is clearly acknowledged.

_____
Signature of Author

DEDICATION PAGE

To my family, for their support.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Online tasks that involve information gathering, those ranging from academic research to vacation planning, often present challenges to users such as information management, clutter and information overload. Studies have shown that users who return to online tasks after an absence have difficulty remembering why particular websites they had saved were useful. This work presents "Webscraps", an innovative web browser extension for Mozilla Firefox, designed to improve information gathering on the Web. Participants in a 30-person user study, significantly preferred Webscraps over webpage "thumbnails" for information gathering tasks that involved comparing information from different websites and remembering important text.

LIST OF ABBREVIATIONS USED

FGS	Faculty of Graduate Studies
FCS	Faculty of Computer Science
Dal	Dalhousie University
WS	Webscraps
PT	Page Thumbnails

ACKNOWLEDGEMENTS

CHAPTER 1   INTRODUCTION

1.1 MOTIVATION AND RESEARCH OVERVIEW

In an information-rich environment such as the Web, tools for information gathering can have a tremendous impact on the productivity of online tasks. Imagine, for example, planning a trip abroad. The prospective vacationer most likely has an idea of places to see, things to do, but the ideas start off vague and are subject to frequent change as planning proceeds. There are also numerous important factors to take into consideration before packing one's passport – flights, hotels, transportation, sightseeing itinerary, food, costs, security, and so on. Each of these in turn has its own sub-category of issues to consider (e.g., hotels: price, location, services/amenities).

Information Gathering is a web task that involves researching a specific topic for such purposes as: comparison, decision-making, item selection, study or idea gathering (e.g., researching and writing a paper or making an online purchase) (Sellen et al., 2002). This work focuses on three subtasks within information gathering: *page revisitation*, *information re-finding* and *information combination*. Page revisitation, the act of returning to a previously visited website has been identified as one of the most frequent tasks in Web browsing (Cockburn et al., 2003). Information re-finding, that is, locating specific information within a page (e.g., words, paragraphs, numbers, figures and tables), involves a variety of cumbersome keeping and re-finding methods (Bruce et al., 2004) to help users successfully re-locate information from previous browsing sessions. Information combination is the subtask of creating a collection of information from several websites "to see the bigger picture" (Amin, 2009). Information combination is also useful for making comparisons between different sources of information.

While tools that assist in page revisitation (e.g., bookmarks and browser history) are supported by most major web browsers, the background research in this study revealed that

very few web browser tools effectively support information re-finding and information combination.

Information re-finding on the Web is often performed using a search engine such as Google (Aula et al., 2005). In such cases, users must rely on retracing the same search path to their previous online destination, which risks being moved or deleted (Obendorf et al., 2007; Teevan, 2005) or forgotten. Information combination subtasks (creating a collection of gathered information) are ineffective when text or word processors are used in combination with web browsers. Once user-selected web page content is copied and stored away from a website in a word processor, it loses the source webpage and its relevant location. Webpage "thumbnails" such as "New Tab Page" in Google Chrome (Google/Inc., n d) and "Top Sites" in Safari (Hiner, 2009), both recent features in these web browsers, support page revisitation but do not address information re-finding nor information combination subtasks.

This work presents *Webscraps* (WS), a tool for information gathering on the Web. Webscraps allows selections of web pages to be created and stored together in a separate tab within a web browser. Webscraps creates a visual "data collage" that allows the user to see and use in one tab all the information is has deemed pertinent, and also enables the user to return to the source page of each item. Results from the user study indicate that participants significantly preferred Webscraps over webpage "thumbnails" for the ability to remember gathered information and effectively compare information.

1.2 RESEARCH PROCESS

A user study was conducted to evaluate Webscraps alongside *Page Thumbnails* (PT), a tool that gathers thumbnails of whole web pages, but unlike Webscraps it does not save specific user-selected areas within web pages. During the study, participants were asked to perform information gathering tasks using each tool. Questionnaires were administered prior and following the tasks, and semi-structured interviews were performed at the

completion of the study. Thirty undergraduate and graduate students from Dalhousie Faculty of Computer Science (FCS) were recruited to ensure a convenient and homogeneous sample. Previous experience using the Mozilla Firefox Web browser was also mandated.

1.3 ORGANIZATIONAL OVERVIEW

This thesis is divided into eight chapters. Chapter 2, *Background Research*, discusses literature relating to Information Gathering. Chapter 3, *Prototype Design*, describes both Webscraps and Page Thumbnails that were developed for the user study. Chapter 4, *Implementation*, provides an overview of the development process for both tools. Chapter 5, *Methodology*, presents the hypotheses and sub-hypotheses, and details the user study, study design and data collection methods used. Chapter 6, *Results*, presents questionnaire and survey results. Chapter 7, *Discussion*, explores the results in detail in order to accept or reject the hypotheses and sub-hypotheses presented in Chapter 5. Finally, Chapter 8, *Conclusion*, summarizes the conclusions from the user study and outlines future work.

CHAPTER 2   BACKGROUND RESEARCH

Background research was concentrated on five areas: (1) Web Tasks and Behaviour, (2) Web Searching, (3) Webpage Revisitation and Re-finding Information, (4) Multi-Session Web Browsing, and (5) Gathering and Managing Information. Each area investigates a different set of user habits, needs and problems that have offered guidance in the design of Webscraps.

2.1 USER BEHAVIOUR

Several studies have looked at Web browsing behaviour. Tauscher and Greenberg (1997b) identified seven Web browsing patterns: (1) First-time visits (accessing a cluster of websites for the first time), (2) Revisits to pages, (3) Authoring of pages (when a user extensively refreshes a page), (4) Using web-based applications, (5) Hub-and-spoke (navigating back and forth between a central page – a "hub" – and several linked pages – "spokes" – listed on the hub), (6) Guided tour (navigating pages that were structured using "Back" and "Forward" links), and (7) Depth-first search (following links deeply from a central page before backtracking).

A client-side long-term study (130,000 page visits) by Weinreich et al. (2008) looked at how people use browser tools when browsing the Web in their everyday environment. Their results suggest that behaviour such as following links or direct access to pages (e.g., entering the address manually in the address bar, retrieving a page from a bookmark, and using the home page button) has remained consistent with what similar studies conducted in the nineties revealed (similar results were found in MacKay and Watters, (2008)). In addition to using bookmarks, Weinreich et al (2008) note the importance of the bookmark toolbar which did not exist in previous research and which, in certain cases of their study, was used solely in place of the bookmark menu.

MacKay and Watters (2008) noted that users tended to store only important bookmarks as opposed to "transitory pages" in the bookmark toolbar. Weinreich et al. (2008) identified a less frequent use of the Back button (from approximately 30% to 15%) due to an increased occurrence of online form submissions (~4% to ~15%) and users opening many more browser windows or tabs (another addition) during a session (~1% to ~10%). These patterns were again apparent in (MacKay and Watters, 2008), where users opened new windows and tabs twice as often as using the Back button.

Researchers have also found that Web users prefer having information immediately accessible (MacKay et al., 2008; Weinreich et al., 2008). This preference means users make use of the bookmarks toolbar instead of the bookmarks menu and keep important pages open in several windows and tabs rather than navigating with the Back and Forward buttons.

2.2 WEB TASKS AND INFORMATION GATHERING

Sellen, Murphy, and Shaw (2002) performed a workplace study to determine what percentage of time users were spending at different tasks (or activities); during 59% of time, users performed "Finding" (specific goal-oriented searches, e.g., looking for recipes or phone numbers) and "Information Gathering" tasks (broader goal-oriented topic researches, e.g., background research to perform comparisons, write documents, or make decisions); "Browsing" occurred in 27% of pages (i.e.: browsing the Web for leisure with no particular goal, e.g., news, magazines, hobbies); "Transactions" (5%), "Housekeeping" (5%), and "Communication" (5%). These results were very similar to the "subtask types" tracked by MacKay and Watters (2008) in an academic user environment.

Kellar, Watters and Shepherd (2007) examined four different information-seeking tasks in everyday Web browsing: "Fact Finding", "Information Gathering", "Browsing" and "Transactions". They found that half of the tasks that users performed were "work- or school-related", while the other half were "personal tasks". Among all the tasks, 46.7%

were Transactions (e.g., checking email, blogging, making online payments), 19.9% were Browsing (e.g., reading news, blogs, game sites and watching or listening to TV/movie and music sites), 18.3% were Fact Finding (e.g., looking up weather facts and finding information for courses and assignments), and 13.4% were Information Gathering tasks (e.g., job hunting, information to make purchases, and finding information for courses and assignments).

Ellis (1993) performed a *paper-based* information-seeking study and identified six categories to represent information-seeking patterns of academic researchers: (1) "Starting" (the initial search activities – e.g., identifying interesting sources), (2) "Chaining" (e.g., "following chains of citations"), (3) "Browsing" (semi-directed searching of an area of interest – e.g., looking through heading or titles), (4) "Differentiating" (filtering information based on its quality and nature), (5) "Monitoring" (monitoring a particular source or group of sources to keep up-to-date with the developments in a particular field), and (6) "Extracting" (systematically looking up information at a specific source). These categories were subsequently used to plot results from a study regarding information seeking on the Web by Choo et al. (2000). Their results indicated that starting and chaining occurred most frequently in "undirected viewing" when a user browsed with no specific goal in mind (Morrison et al. (2001) describe this behaviour as "Exploring"). Differentiating, browsing and monitoring occurred more commonly in "conditioned viewing" (viewing selected topics or certain kinds of information), differentiating and localized extracting in "informal searches" (researching to gain knowledge on a specific topic), and extracting occurred the most in "formal searches" (looking for specific information for a predefined topic) (Ellis, 1993).

2.3 WEB SEARCH BEHAVIOUR

Web search behaviour is an important step in information gathering. White and Drucker (2007) studied Web searching behaviour of over 2500 users and identified two classes of interaction variance (searching behaviour): (1) a low variance model called "Navigators"

and, (2) a high variance model called "Explorers". Navigators exhibited "consistent interaction patterns in the [search] trails they follow[ed]", meaning they often started with a single search query and navigated links sequentially toward their problem resolution. Explorers on the other hand, were characterized by their higher variance searching patterns, performing more queries (more branching) and relying on multiple strategies during a particular search. Navigator-style behaviour occurred more frequently in "well-defined fact-finding tasks" but Navigators were more likely to revisit domains than their Explorer counterparts. White and Drucker also found that 12.5% of websites were part of a "search trail" while other websites were concerned with email, e-commerce or home page browsing. Interestingly, during their searches users spent more time manually browsing for information (71% of websites) by manner of clicking on the immediate links of a website versus performing queries using a popular search engine (29% of websites).

2.4 INFORMATION GATHERING FACTORS

2.4.1 Page Revisitation

There are several studies that investigate *Page Revisitation* on the Web. Tauscher and Greenberg (1997a) characterized web browsing as a "recurrent system", finding a 58% percent probability that a user's next web page will be a revisit. Their participants revisited pages in situations when information had changed, in cases where they wanted to explore pages further, or when "the page [was] on a path to another revisited page".

A more recent investigation of revisitation patterns by Adar et al. (2009) found that users were drawn to frequently updating pages primarily to find or monitor new information. In an earlier study (2008), Adar et al. clustered revisitation patterns into four "cluster groups" based on their intervals of revisitation: "Fast Group" (intervals of less than an hour), "Medium Group" (hourly revisits), "Slow Group" (intervals longer than a day), and "Hybrid Group" (intervals less than an hour or longer than a day). They found that

different tasks displayed different revisitation intervals. For example, shopping was generally in the fast group, whereas viewing personal data occurred most frequently in the slow group. They also noted that "as the revisitation period increased, the pages were more likely to be accessed from a search engine".

An important reason for website revisitation is *Multi-Session Web browsing*, where more than a single session is required to complete a task. MacKay and Watters (2008) found that approximately a third of web sessions belong to a multi-session, where users spend an average of 2.61 sessions working on a particular task. In their study, participants frequently made use of other applications such as word processors or text editors in order to compare or store information for their task.

Other researchers noted similar behaviour. Sellen, Murphy and Shaw (2002) demonstrated that information gathering was hampered even on large screen devices when tasks required more than a session to complete. Users frequently resigned to copying information into notebooks, printing, and collecting printouts in folders to help manage their task.

2.4.2   Re-finding Information

Information re-finding is characterized by the act of returning to a particular area in a web page to locate specific information (MacKay et al., 2005; Teevan, 2005). Bruce, Jones and Dumais (2004) noted that users were 93% successful at information re-finding using the first of four "first-try" methods: (1) entering the URL directly, (2) accessing Favourites or Bookmarks, (3) using a search engine to perform a search, and (4) accessing the link using another website. In 76% of successful re-finds, no keeping methods were used (e.g., methods 1, 3, and 4). Users were generally able to re-find specific information in less than a minute.

2.4.3 Gathering and Managing Information

Users create and manage personal information "spaces" to overcome certain complexities of the Web. For instance, Abrams and Baecker (1997) noted that users created personal archives, such as a collection of bookmarks, to "prevent information overload" in a medium that has "no aggregate structure" with "redundant, erroneous and low quality information". They described these efforts as "trying to reduce entropy through maintenance", "add[ing] structure" and "compensating for the lack of a global view by creating their own personal view".

Bookmarks, as described by Abrams et al. (1998), provide three major uses: (1) reducing user load, (2) facilitating navigation and access, and (3) collaborating, publishing and archiving. Abrams et al. found that, while participants appreciated bookmarks for their "mnemonic" properties, saving them from manually typing in (or remembering) URLs, problems arose when users tried to manage larger collections of bookmarks over time. For example, participants had difficulty remembering the purpose of a stored bookmark or determining whether it was part of a single or multi-session web task. Others found that when trying to organize bookmarks, they lost the "time-ordered view" in the menu that helped them remember each bookmark's purpose. Participants also noted that managing a smaller collection of bookmarks allowed them to see the entire bookmark menu on a single screen, reducing the effort required to locate previously viewed information.

In a study of Personal Information Management (PIM) using bookmarks, emails and files (Boardman and Sasse, 2004), users reported that PIM was an important area frequently causing problems and frustration. Bookmarks were, however, found to be less valued than other collections such as emails and files, and tended to be much smaller (their numbers averaging in the tens whereas file and email collections averaged in the thousands). Reasons for their avoidance included (1) "the existence of other ways of re-accessing websites" such as performing search queries, and (2) the idea that URLs change frequently and create revisitation problems when the address stored in the bookmark is no

longer valid. For these reasons, users preferred saving documents from the Web locally rather than bookmarking them.

Abrams et al. (1998) observed four habits of bookmark organization: (1) "no filers" (participants that never organized bookmarks); (2) "Creation-time Filers" (participants that stored bookmarks and immediately placed them into a proper folder); (3) "End-of-session Filers" (participants that organized their bookmarks at the end of a session); and (4) "Sporadic Filers" (participants that occasionally organized bookmarks or did so according to a schedule). Users with larger bookmark collections (e.g., > 300) tended to file bookmarks immediately into proper categories "in order to manage so many bookmarks" (Abrams et al., 1998). For such users, "maintaining an up-to-date organizational structure outweigh[ed] the distraction from the browsing that is required to file a bookmark". These results are consistent with earlier work from Whittaker and Sidner (1996) into strategies for handling email overload. In a later study by Boardman and Sasse (2004), participants described filing a subset of bookmarks while leaving unsorted bookmarks as reminders. Abrams et al. (1998) also identified seven different methods by which their participants organized bookmarks (Table 1).

Table 1: Methods for Organizing Bookmarks (adapted from Abrams *et al.* (1998))

Organization Method	Description
No organization	Bookmarks stay in the order created
Ordered List	Users manually re-arrange a list
Set	Users create folders to categorize bookmarks
Hierarchy	Users create folders within folders
External	Users export bookmarks to a separate program
Web Page	Users create Web pages out of their bookmarks
Other	Users search, sort and use unusual methods

Research to help organize information to make it more easily navigable and accessible has been around well before the Web era. Malone's classical study (1983) identifies two

types of paper management techniques: "filing" and "piling" and has since been applied to Web studies (Boardman and Sasse, 2004; White and Drucker, 2007). Filers tend to file papers, keeping an organized desktop, whereas pilers tend to keep "loosely characterized piles" of paper on the desktop. Malone found that piling was caused by (1) the overall difficulty of organizing files (e.g., labelling folders or binders), (2) the cognitive load of creating categories and classifying information for easy retrieval, (3) the need to keep "to do list" items on the surface, and (4) keeping information that is frequently used easily accessible. Unfortunately, as piles grow, retrieval becomes a more challenging task (Malone, 1983). While in Malone's time it was widely believed that filing was a more superior habit than piling (Malone, 1983), Whittaker et al. (2001), found that during their study of paper-based PIM in an office environment, filers in fact amassed more superfluous information than pilers by means of "premature filing" – filing papers regardless of their usefulness in order to keep the desktop clean. Filers also had greater difficulty re-accessing papers in their collections, as categories were forgotten over time, causing instances of duplicate filing (Whittaker and Hirschberg, 2001).

Regardless of the medium in which they work, users can often be pilers or filers. Either information management method presents its own problems and advantages, as noted by Whittaker and Hirschberg (2001). On the Web, while bookmarks have long been a method for organizing information (Abrams and Baecker, 1997), larger bookmark collections require maintenance as they grow to overcome structural complexity (Abrams et al., 1998).

2.4.4   Other Factors for Information Gathering

Findings by Sawasdichai and Poggenpohl (2002) reported frustration among users when web pages contained too much irrelevant visual or textual information, or failed to present "better-categorized retrieval results" from built-in searches. During tasks that required decision-making (e.g., making a purchase), "novice" users who were less familiar with a web page and its content made decisions toward the end of their session after spending most of their time searching, gathering information and browsing web pages.

"Expert" users familiar with a web page and its content performed decision-making all-throughout their session. Certain users employed different search strategies but were often led to the same pages and became frustrated. Other participants either gave up searching altogether or sought help from more experienced users.

2.5 WEB TOOLS FOR GATHERING

Several tools and recommendations have been suggested to improve information gathering tasks. They are summarized in the following paragraphs.

Sawasdichai and Poggenpohl (2002) proposed that tools that assist in decision-making take into account three important stages: *early* – eliminating irrelevant results, *comparative* – finding the most interesting results, and *final* – comparing the details of several results.

Sellen, Murphy and Shaw (2002), identified a need for a type of "dynamic Webscrapbook" that could "save information not just in terms of URLs, but also in terms of a much wider range of information" (e.g., text, graphics and search results). They envisioned such scrapbooks as an opportunity to avoid the use of hierarchical folder structures such as bookmarks. They suggested that this model is "for the most part, entirely unsuitable for supporting [sic] on small screen devices in mobile settings."

Other studies have focused on tools to improve page revisitation and information re-finding. MacKay and Watters (2005) proposed "Landmarks", a bookmark extension allowing users to return to specific areas of web pages. Their study found that participants were able to re-find facts significantly faster, especially when information was located in areas away from the beginning or end of a webpage.

Similar capabilities, "scrollbar marks" and "Footprints" (a scrollbar with marks and mark thumbnails), were proposed by Alexander et al. (2009) to help users return to previously

viewed areas in a document. Their study showed that users frequently revisited certain "document regions". They reasoned that a short list of frequent locations, such as the scrollbar marks provided by their model, could make re-finding more efficient in most scenarios.

Finally, Robertson et al. (1998) developed "Data Mountain" to visually arrange document thumbnails in 3D using a 2D interaction technique. In their study, they found that users formed spatial relationships between files using spatial memory and were able to store and retrieve bookmarks more quickly than using traditional bookmarking management methods.

CHAPTER 3   PROTOTYPE DESIGN

3.1 WEBSCRAPS – A TOOL TO MANAGE WEB INFORMATION GATHERING
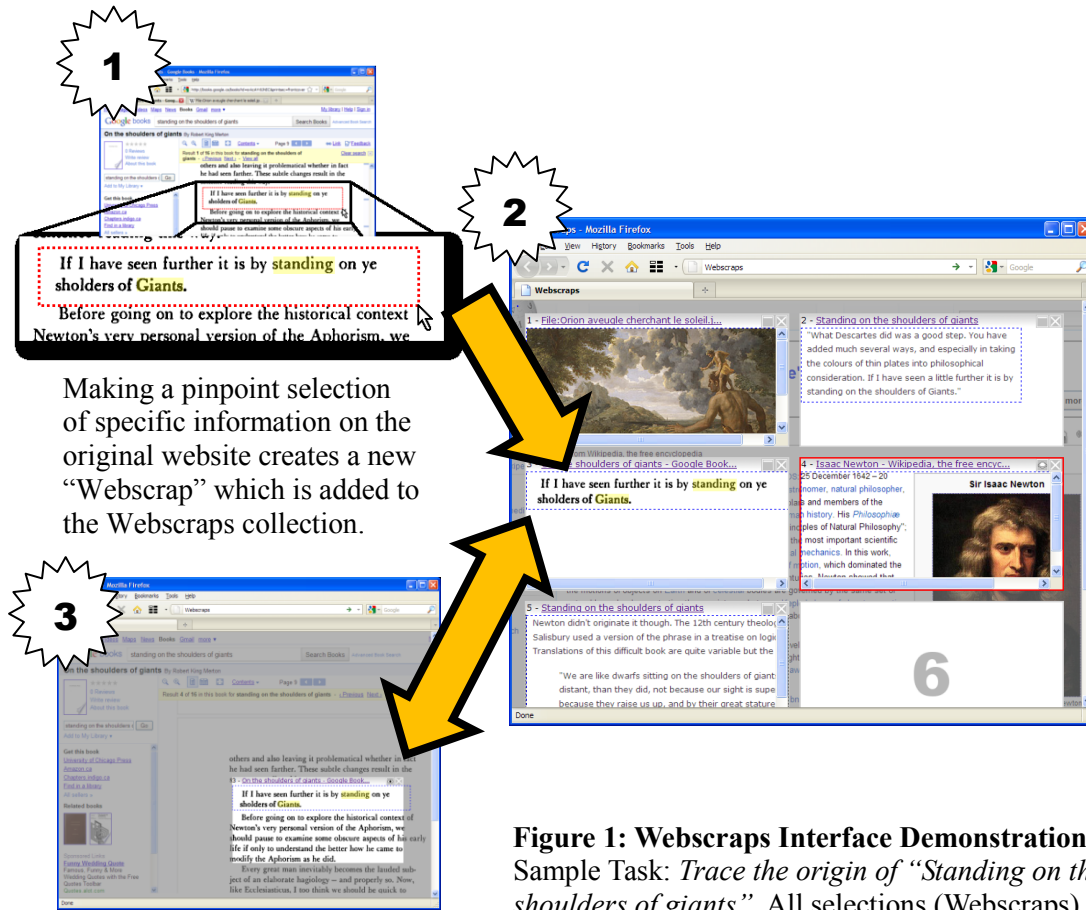
Making a pinpoint selection of specific information on the original website creates a new "Webscrap" which is added to the Webscraps collection.

One click on any Webscrap returns the user to the specific area within the original webpage for more information.

Figure 1: Webscraps Interface Demonstration
Sample Task: *Trace the origin of "Standing on the shoulders of giants".* All selections (Webscraps) come from different areas of one or many websites. A single "local" Webscrap (Newton's photo above) also shows its context in the shaded background behind other Webscraps. All other Webscraps are "remote" but each in turn can be made local according to the user's preference.

Using the findings discussed in the previous section, the Webscraps tool (Figure 1) was developed to help perform page revisitation, information re-finding and information combination subtasks live within a web browser interface. A second prototype mimicking a recent addition to browsers, "Page Thumbnails" (Figure 2), was also developed.

Both tools were developed as Mozilla Firefox extensions for the user study over a six month period using JavaScript and Mozilla's XML User Interface Language (XUL).
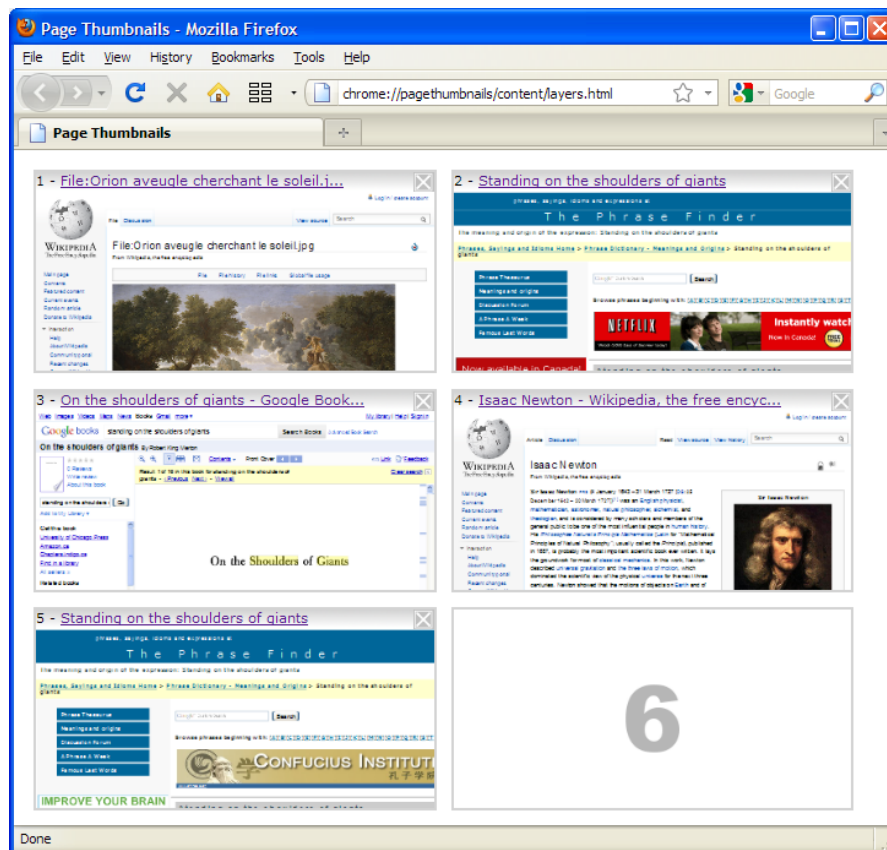
Figure 2: Page Thumbnails, a second prototype used during the study, mimics the reduced features presently available in Google Chrome and Apple Safari. Unlike Webscraps, Page Thumbnails does not save user-selected areas within web pages.

3.1.1 Webscraps Features

Webscraps introduces a gathering entity called the "Webscrap", an area selected by the user from any webpage using a pointing device such as the mouse. A menu activates the tool, enabling selection of a region of interest (see Figure 1, top callout) within any open webpage (e.g., words, paragraphs, numbers, figures or tables). The Webscrap is then saved in the Webscraps interface, a dedicated tab in the web browser. The Webscraps interface allows selections from several web pages to be combined into a single view. Importantly, this approach facilitates re-finding specific information without requiring a

revisit to the original website. Furthermore, since each Webscrap appears next to other Webscraps in the same collection, the user can easily perform direct comparisons between each Webscrap within a single web browser page tab.

In the event that a user wishes to return to the original webpage of a Webscrap for additional information, a button on the specific Webscrap opens up the original location on the source website (see Figure 1, bottom callout). Webscraps has two visibility modes: *local* and *remote* (Figure 3). A local Webscrap displays its original website as a shaded background – the Webscraps "Layer" – behind all other Webscraps in the collection. All other Webscraps are remote, displaying just the selection without a contextual background, but each in turn can be made local according to the user's preference.
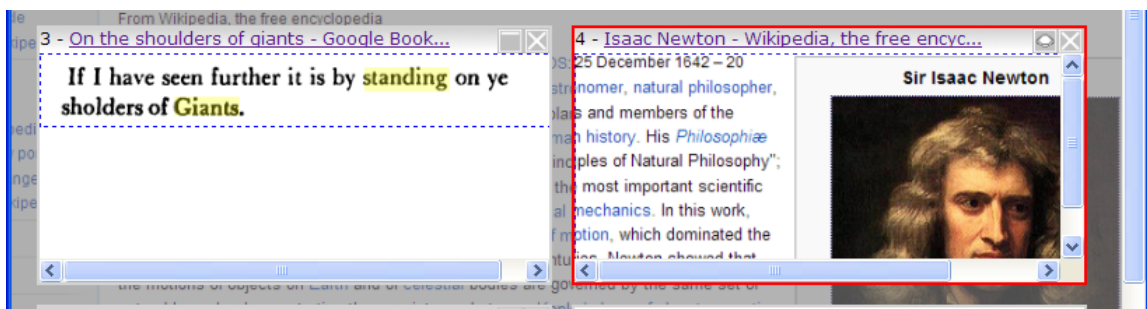
Figure 3: A remote Webscrap (left) and local Webscrap (right); the local Webscrap displays its original website as a shaded background behind all other Webscraps in the collection.

3.1.2 Page Thumbnails

A second prototype, Page Thumbnails (Figure 2), was developed for the user study as a comparison to the Webscraps tool. Page Thumbnails was designed to mimic features in certain web browsers (e.g., "New Tab Page" in Google Chrome (Google/Inc., n d) and "Top Sites" in Safari (Hiner, 2009)). Page Thumbnails captures a screen shot of an open page and displays the thumbnail collection in a dedicated tab in a manner similar to Webscraps. The original website can be reopened by clicking on a thumbnail. Unlike Webscraps, Page Thumbnails does not allow pin-point selections and does not save specific information located by the user on web pages.

3.2 WEBSCRAPS FEATURES: FROM RESEARCH TO IMPLEMENTATION

In this section, features are discussed that have been implemented in Webscraps as a result of guidelines identified in the Background Research chapter.

3.2.1 Information Gathering

Information gathering is an important task that users frequently engage in both personal, academic and workplace environments (MacKay and Watters, 2008; Sawasdichai and Poggenpohl, 2002), and which requires more time to complete than other Web tasks (Kellar et al., 2007; Sellen et al., 2002). Yet, several studies mention a lack of tools to support information gathering in different situations (Adar, Teevan, and S. T. Dumais, 2009; MacKay and Watters, 2008; Sawasdichai and Poggenpohl, 2002; White and Drucker, 2007). Others have suggested innovations to help improve information gathering (MacKay et al., 2005; Sellen et al., 2002).

Webscraps focuses on improving three *subtasks* (hereafter referred to as tasks) of information gathering by:

- providing a means to gather and re-find *specific information* on websites,
- providing effective *page revisitation*, and
- allowing related data to be *combined* / *"clustered"*

3.2.2 Information Re-finding and Page Revisitation

Problems with traditional search: While a few studies have shown that searching the Web (Aula et al., 2005; White and Drucker, 2007) or searching local bookmarks and cache (W Jones, H Bruce, and S Dumais, 2003; Kellar, Watters, and Shepherd, 2007) are popular and effective methods of page revisitation, searching to re-find information re-

quires remembering the specific search query and finding the location of the page within the search results (Teevan, 2005). For re-finding several sources of information, such repeated search procedures quickly becomes inefficient.

Problems with "copy and paste": Users often store information in word processors or text editors between tasks (MacKay and Watters, 2008). Unfortunately, once user-selected web page content is copied and stored away from a website in a word processor, it loses the source webpage and its relevant location. If the source URL of the copied information has been saved, the area of interest must still be located within the page upon revisitation. If the URL is missing, an additional searching procedure is necessary to locate the website. Furthermore, copying and pasting is an additional step that requires transferring information from one location to another. Whether copying to email or to an outside text editor, the user is responsible for remembering where information is saved (Jones et al., 2001).

Re-finding and Revisitation in Webscraps: Webscraps draws on guidelines proposed by White and Drucker (2007) for a tool that "jump[s] directly to [the] information target, with no steps in-between" when performing "short, directed search trails" (see also: "Teleporting" in (Teevan et al., 2004)). Webscraps are selections of information that are linked to their source web pages and are located in a dedicated browser tab for quick reference. As opposed to bookmarking or using web page thumbnails, Webscraps does not employ mnemonic devices (e.g., titles, images) to remember gathered information. Instead, selected information is directly displayed in the Webscraps interface. Revisiting a web page for more information is performed by clicking on a selection within Webscraps to reveal the page and location from which it was gathered. Additionally, local Webscraps display their background behind all other selections (Figure 3) in order to provide a context to assist the user to recognize selections.

3.2.3 Combining Information

The problem of combining related information gathered online has been the topic of a few studies (Robertson et al., 1998; Sellen et al., 2002; Sugiura and Koseki, 1998). Webscraps draws upon suggestions made by Sellen et al. (2002) for a type of mixed-media scrapbook for information gathered online:

> *"The need to save information not just in terms of URLs, but also in terms of a much wider range of information (selected pieces of text, search results, graphics, etc.), implies that more flexible ways of managing information would be valuable. Rather than saving to hierarchical folders, users could construct mixed-media scrapbooks into which they clip and store information from Web pages for particular projects."*

By interesting coincidence, Sellen et al. chose the name "Webscrapbooks" to describe these recommendations.

Combining information in Webscraps: Webscraps stores selected information within a dedicated browser tab instead of using a traditional folder/menu layout commonly used with bookmarks. Storing Webscraps side-by-side within a browser window provides opportunities for comparison of data gathered from different web pages or different locations of the same web page.

CHAPTER 4   IMPLEMENTATION

This chapter discusses the design, mock-up and implementation phases of Webscraps and Page Thumbnails. Content in this chapter was adapted from a final course project report for CSCI6406.

4.1 DESIGN AND MOCK-UP

Before extension implementation started, a mock-up interface was sketched (Figure 4) and built using static HTML and Cascading Style Sheets (CSS) (Figure 5). The mock-ups provided an opportunity to plan the spatial arrangement of interface components (e.g., selections, links, buttons, and icons) and determine whether any additional interface controls were necessary. Both Webscraps and Page Thumbnails were designed to share similar layouts in order to minimize the effects of "unimportant" differences in usability in (e.g., opening up Webscraps/Page Thumbnails) and to allow the study to focus on the important features of both tools (e.g., making selections / saving web pages). The initial 6-box layout scheme for Webscraps and Page Thumbnails as they appear in the browser viewport is shown in Figure 4. Boxes represent web page selections in Webscraps and thumbnails of pages in Page Thumbnails.

Screen Resolution: The interfaces were designed to allow two columns of boxes to fit horizontally and three to fit vertically and to avoid the need to scroll within a 1024x768 (4:3) resolution browser viewport. Although recent LCD/LED monitors generally support higher resolutions, the 1024x768 resolution was adopted as a minimum resolution. The interface is supported on higher resolution displays. During the user study, a 1280x1024 (5:4) screen resolution (non-widescreen) was used.

Number of boxes: For the purposes of the user study, the number of Webscrap and Page Thumbnails boxes was chosen to match the number of items to be gathered during each

task. Additional or an unlimited number of boxes would be available for everyday use. Part of the study involved asking users to suggest ways to manage the number of boxes in Webscraps and Page Thumbnails.
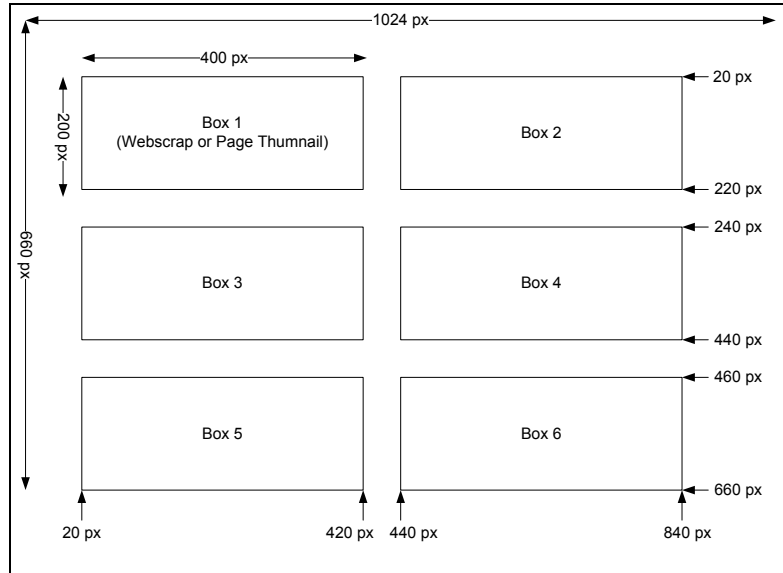
Figure 4: Layout of the Webscraps and Page Thumbnails Interfaces (Within Web Browser Viewport)
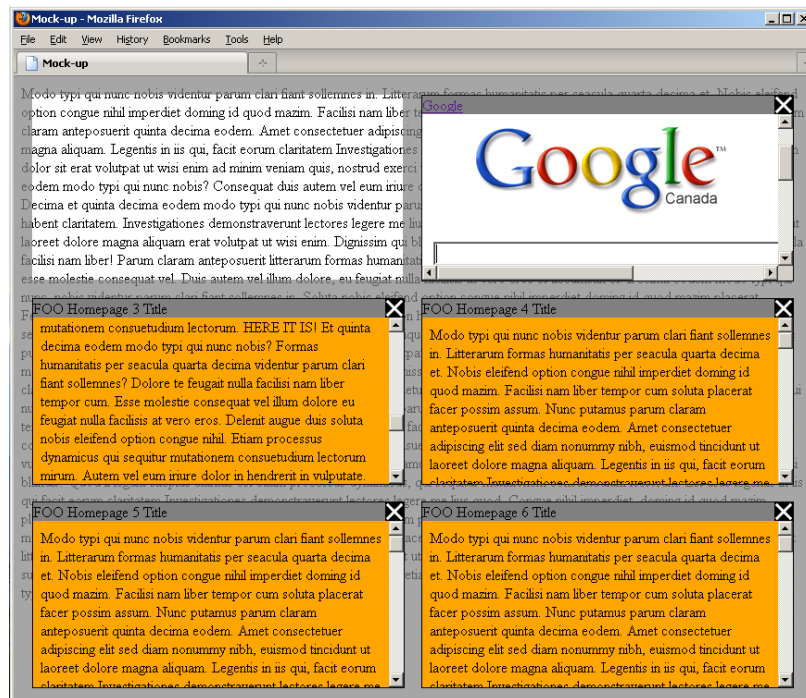
Figure 5: A Webscrap mock-up showing local and remote scraps, and the surrounding local scrap layer

4.2 Firefox Extension Development

4.2.1 Introduction

Over the course of six months, technologies for Firefox extension development such as JavaScript and Mozilla's XML User Interface Language (XUL) were studied. The Mozilla Development Centre website[1] (MDC) provided helpful resources such as tutorials, sample source code, references to other extension projects, and updates of recent innovations and browser compatibility changes. Frequent reference was also made to "JavaScript, The Definite Guide" by Flanagan (2006) and online forums discussing different components of development under Mozilla's architecture.

A Different Style of JavaScript: In Firefox development, JavaScript is used as the primarily backend language, although in a different role than traditional client-side JavaScript. Client-side JavaScript is usually bound to a single website and has limited operating system privileges. JavaScript used in Firefox extension development remains active within the web browser across all pages visited and contains greater operating system privileges such as file access and the ability to capture screen shots. These differences require that extension code be robust across many different web pages. For example, code that accesses the document object model (DOM) must be sufficiently generic and robust to be compatible with many web pages. With increased system privileges, secure JavaScript code becomes especially important. This different style of JavaScript programming produced specific challenges and code bugs throughout development that were not typically addressed by standard JavaScript documentation. The MDC website provided helpful forums, coding tips and examples to help with these specific issues.

---

[1] developer.mozilla.org

4.2.2   Structure and Technologies of Webscraps and Page Thumbnails

Firefox extensions follow a standard Mozilla file and folder hierarchy (Figure 6). By convention, the top-level folder is given the extension name, followed by an "at" symbol (@) and the developer/business domain name. The mandatory folders "chrome" and "content" contain configuration and source code files.
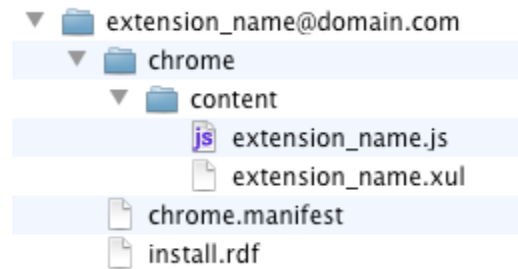
Figure 6: Basic Folder Layout of a Firefox Extension

A minimum of four files are required in a Firefox extension: (1) `chrome.manifest`, (2) `install.rdf`, (3) `extension_name.xul`, and (4) `extension_name.js`. The file `chrome.manifest` keeps track of files that modify Firefox's "Chrome" graphical user interface (e.g., `extension_name.xul`). `install.rdf` provides information about the extension (e.g., extension name, author name, compatible Firefox versions). `extension_name.xul` and `extension_name.js` are source code files for adding user interface controls and for backend programming, respectively. Other technologies such as HTML and CSS may also be used as part of the extension interface but each file must be contained within the extension's folder hierarchy.

For distribution and installation purposes, the folders are packaged as a single ZIP (archive) with the filename extension renamed to "XPI" (representing "Cross-Platform Installer Module"). During installation, the archive is unpackaged to the user's browser profile directory. A separate Firefox profile was used during development as a safeguard against possible corruption of personal browsing data.

XUL: Any user interface modifications that a developer wishes to make to Firefox's native XUL interface (known as the default Mozilla Chrome interface) such as the global mouse menu, custom toolbars and buttons, must be defined in `exten-sion_name.xul`. This procedure is known as "overlaying" the default browser interface. For Webscraps and Page Thumbnails, a menu button was added to the main Firefox toolbar to provide access to the interfaces (Figure 7). Pressing the left "main" region of the button (Figure 7, left) opens up Page Thumbnails or Webscraps in a new browser tab. Pressing the right region of the button (Figure 7, right) opens a drop-down menu (Figure 8) allowing the user to initiate a page selection or saving procedure.
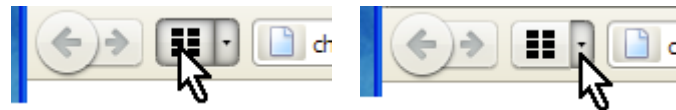
Figure 7: Opening the Webscraps/Page Thumbnails Tab (Left); Opening the Menu (Right)
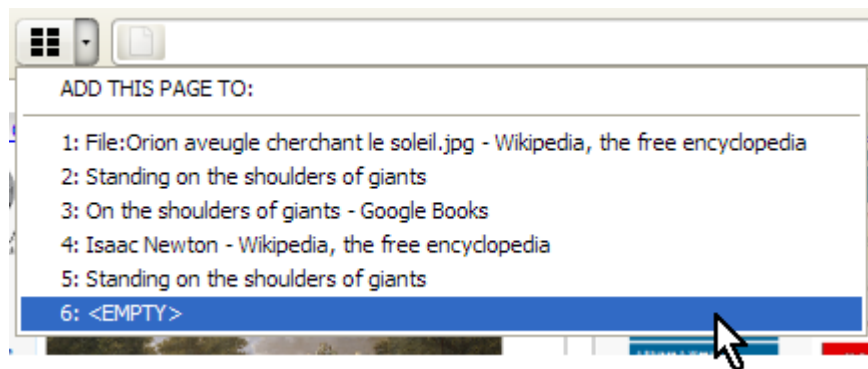
Figure 8: Menu in Webscraps and Page Thumbnails for saving information

Source code written in JavaScript (e.g., event handler functions and objects) is stored in `extension_name.js` and is executed when Firefox loads. Events from XUL interface controls are passed on to event handler methods in JavaScript in a similar manner as Forms and Input are handled in HTML.

URI : A convenient Universal Resource Identifier (URI) provides operating system independent access to each of the extension files. For example, the following URI references the interface file for Webscraps:

```
chrome://webscraps/content/webscraps.xul
```

URIs were used to reference source code and bitmap files during development.

XPCOM: Mozilla uses a Cross Platform Component Object Model (XPCOM) to provide advanced functionality such as file and memory management. XPCOM was primarily used during development for file access operations within the browser profile and for making screen shots of web pages (e.g., saving thumbnails of web pages). Development using URI and XPCOM enables Firefox extensions to run in any operating system that supports Firefox.

Object Serialization: Mozilla provides convenient object serialization by means of a native implementation of JavaScript Object Notation (JSON). Variables within objects are exported using the `JSON.stringify()` function and parsed using `JSON.parse()`. In Webscraps and Page Thumbnails, gathered data such as hyperlinks and page titles are saved between browsing sessions using JSON.

4.3 PHASES OF IMPLEMENTATION AND CHALLENGES

Development of Webscraps and Page Thumbnails was divided into several milestones, based on the major features being developed in each tool. Development was performed under Windows and MacOSX using the NetBeans integrated development environment. Development started on Firefox 3.5 and continued on version 3.6. Debugging was performed with the help of the Firefox "Firebug" extension[2].

---

[2] Available from: https://addons.mozilla.org/en-US/firefox/addon/firebug/

4.3.1 Setting-up the Folder Hierarchy (Both Tools)

During the initial phase of development, the necessary extension directories and source code files were created and registered as a Firefox extension. To test that all elements were properly registered, a new browser button was created using XUL and linked to a JavaScript event handler method that displayed a conventional programmer's greeting ("Hello World") upon clicking. The appearance of "Hello World" signalled that all files were properly registered, and thus development of features could begin. This milestone required a significant period of time to complete as several attempts were made before all files were successfully registered.

4.3.2 Mouse Selection (Webscraps)

To gather selection coordinates and visualise the selection procedure, event handler methods were implemented in JavaScript. Three methods handle mouse selection by listening to signals from the mouse called "events": (1) startSelectionBox, (2) updateSelectionBox, and (3) finishSelectionBox. The initial XY page coordinates are recorded in (1) by listening to the onmouse**down** window event. During a short period while the mouse button is pressed and the mouse is being dragged, mouse movement is captured and co-ordinates are updated by listening to onmouse**move**. The final height and width attributes are captured when onmouse**up** occurs and the final coordinates are passed on to the script. Four variables, scrapH (height), scrapW (width), scrapX, and scrapY, are used by other functions to save scraps and later position them within the interface for viewing. These are saved using JASON serialization. Visualization of the selection procedure was implemented by drawing an HTML DIV box over the area being selected within the page. Drawing the selection box involves temporarily inserting a DIV element into the web page being selected.

4.3.3   Saving and Rendering Webscraps and Page Thumbnails

The most interesting challenge faced during both Webscraps and Page Thumbnails development was how to save and render page selections (Webscraps) and miniature web page thumbnails (Page Thumbnails).

Problems encountered while using iFrames: During initial development, iFrames were used to display sections of web pages as Webscraps. A host HTML page contained several iFrames, each displaying a remote website's selected area positioned within view (Figure 5). Unfortunately no solution was found that would allow a user to automatically scroll to specific positions within iFrames. Page layout problems were also encountered when rendering websites that used different combinations of "absolute" or "relative" positioning styles as these caused page contents to shift within in the iFrame window, frequently occluding the content of interest.

The Canvas Tag: It was eventually discovered that the canvas HTML tag provided a solution for saving and rendering Webscraps and Page Thumbnails. A lesser known feature of the canvas tag is its ability to render entire areas or subsets of web pages as bitmap images. This feature is not available in client-side JavaScript but is available in JavaScript for extension development purposes. In addition to capturing pages, the canvas tag also supports cropping and scaling of images. Cropping was used in conjunction with mouse coordinates from selection steps to render specific web page areas for Webscraps. Scaling was used to shrink browser view port images into Page Thumbnails. Images from the canvas tag were exported as PNG files to the profile directory and re-read by the extension at loading time.

4.3.4   Main Interfaces

The last major milestone involved implementing the main interfaces of Webscraps and Page Thumbnails. A template HTML page was stored within the extension directory and was used to build the Webscraps and Page Thumbnails interfaces at runtime. At first run,

an unmodified template provided the initial appearance of both Webscraps and Page Thumbnails (Figure 9). As Webscraps or Page Thumbnails are gathered using the toolbar menu button by a user, they are inserted via JavaScript into the empty boxes along with the title and link of the source page. Additional interface controls (Table 2) are also added to the boxes.
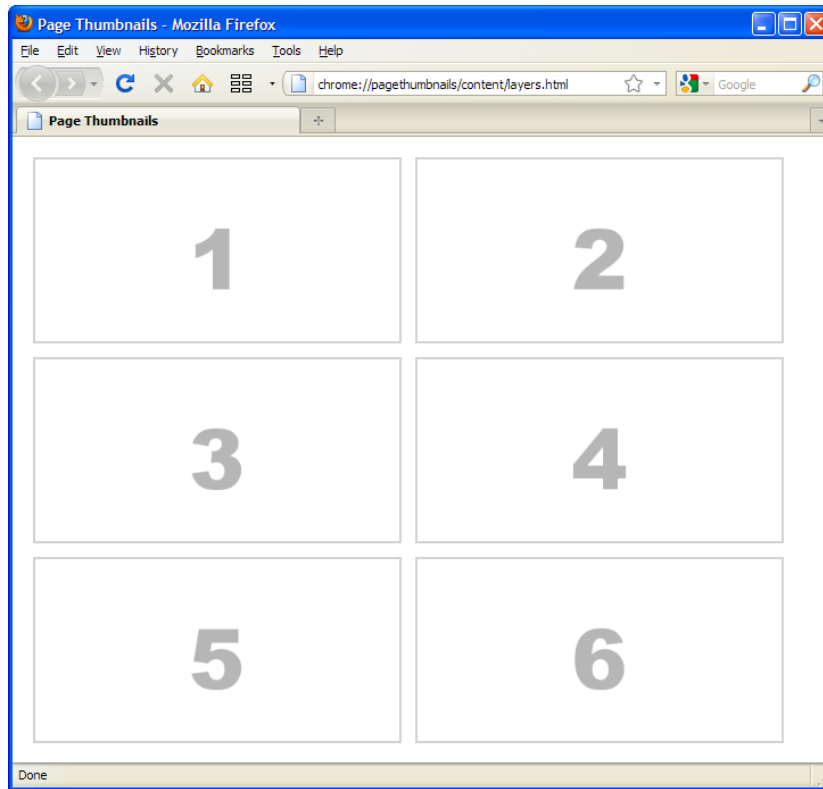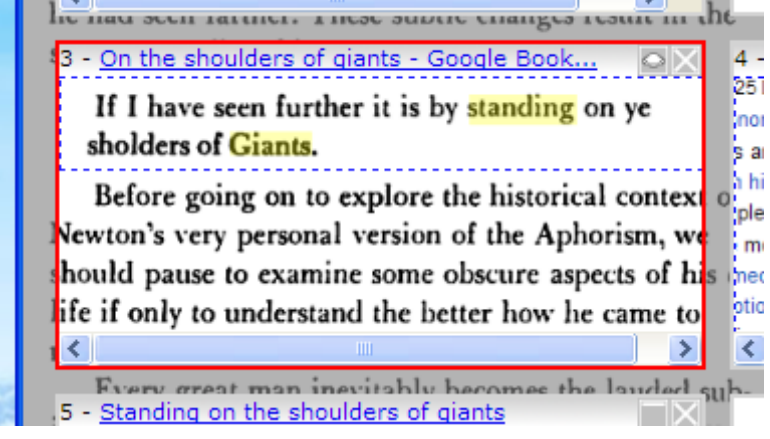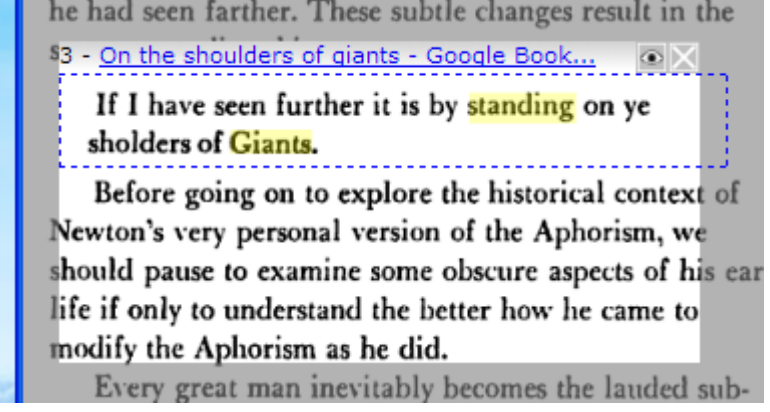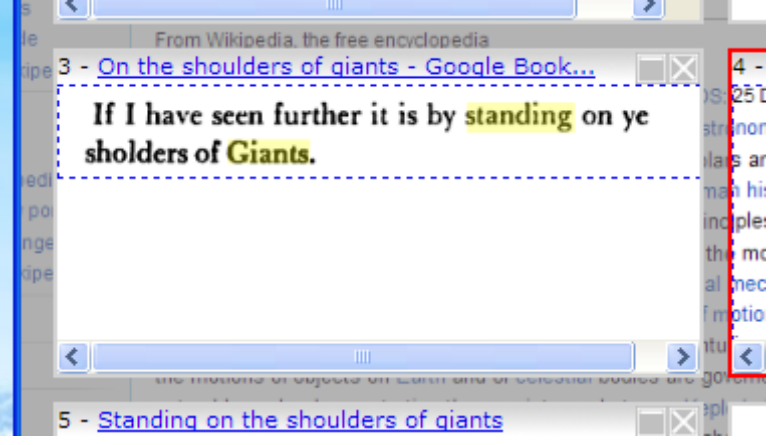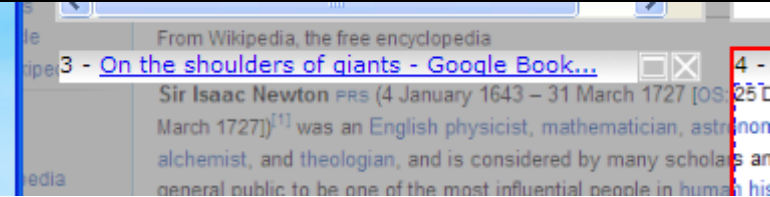
Figure 9: Initial Layout of Webscraps and Page Thumbnails

Table 2: Webscraps and Page Thumbnails User Interface Controls
(* = Webscraps Only)

Icon	Name	Description
	Delete Box	Deletes the current Webscrap / Page Thumbnail
	Minimize Box[*]	Minimizes the current Webscrap
	Restore Box[*]	Restores the current Webscrap
	Open Layer Peek[*]	Hides all remote Webscraps
	Close Layer Peek[*]	Restores all remote Webscraps

Table 3: Different Viewing Modes of Local and Remote Webscraps

A. Local Webscrap / *Normal*	
B. Local Webscrap / *Layer Peek On*	
C. Remote Webscrap / *Normal*	
D. Remote Webscrap / *Minimized*	

Webscraps have four different viewing modes (Table 3). Each of these viewing modes has a title bar that contains the title of the source page as a hyperlink (to open the page in a new window) along with different user interface controls.

Local Webscrap / Normal: When a Webscrap is added to the interface, it is initially represented as a local Webscrap (Table 3A). The background behind all Webscraps (or any remaining empty boxes), the *Layer*, is changed to the web page of the local Webscrap. This is achieved by inserting a large bitmap of the source web page behind all the boxes and lining it up against the local Webscrap at the position from which the local Webscrap was gathered. Each Webscrap (local and remote) is also surrounded by a DIV with a dashed border to mark the area where selection was performed. Boxes containing Webscraps are fixed in size but provide scrollbars if their area is exceeded by a Webscrap.

Local Webscrap / Layer Peek On: The Webscraps interface provides a feature called "Layer Peek" that hides all other Webscraps for when a user wishes to browse the source web page for more information (Table 3B). Layer Peek is activated by pressing the Open Layer Peek button on its title bar and disabled by the Close Layer Peek button (Table 2). Hiding of boxes was achieved by setting their CSS visibility style to "hidden". Other regions of the source web page beyond the browser viewport can be viewed by dragging the Layer with the mouse.

Remote Webscrap / Normal: At any point, only one Webscrap is in the local viewing configuration while others are remote (Table 3C). Remote Webscraps do not show their contextual background (Layer), but can be transformed into local Webscraps by double clicking on their title bar.

Remote Webscrap / Minimized: Remote Webscraps can also be minimized to increase the viewing area behind them. Minimizing is achieved by clicking on the Minimize Box button on the box's title bar, and restoring is achieved by clicking on the Restore Box button (Table 2).

Deleting Webscraps and Page Thumbnails: Webscraps and Page Thumbnails can be deleted by pressing the Delete Box button (Table 2) in the title bar. During testing among students prior to the user study, it was determined that a confirmation dialog box needed to be implemented to avoid unintentional deletions.

Viewing Page Thumbnails: Page Thumbnails offer a single viewing mode (Figure 10). Like Webscraps, the title bar contains a link to open the web page in a new window and a button for deletion.

Figure 10: Viewing a Page Thumbnail

4.3.5   Conclusion and Future Work

Building Webscraps and Page Thumbnails was a challenging development experience. JavaScript code in both extensions needed to be robust to function with many different web pages. By contrast, client-side JavaScript that runs on a website needs to conform to a handful of different web browsers.

During the implementation phase, a choice was made between rendering Webscraps and Page Thumbnails live from a web page (e.g., using iFrames) or using screen shots of web pages. Live information provides advantages such as the ability to select and search through text, however, implementation of this feature is not trivial. In addition, as web pages are updated, information already gathered (e.g., selected areas) risks being lost.

Bitmaps are not affected by updating as the gathered information is preserved locally on the user's machine as a screen shot.

An intermediate approach is web page caching. Caching, however, would require considerably more implementation and would still not overcome other rendering difficulties such as embedded video and possibly displaying dynamic web pages (e.g., Facebook).

Future versions of Webscraps and Page Thumbnails will introduce more features such as the ability to save information using more than six boxes. Feedback from the user study will also be considered. Finally, other devices (e.g., iPads) and web browsers will be considered in future development.

CHAPTER 5   METHODOLOGY

5.1 HYPOTHESES AND RESEARCH QUESTIONS

The following hypotheses and sub-hypotheses are considered in this work:

H$_1$ Webscraps is a more *effective* gathering tool than Page Thumbnails for information gathering tasks.

> **H$_{1.1}$**   Webscraps increases users' ability to remember gathered information.
>
> **H$_{1.2}$**   Webscraps helps users more effectively recognize websites.
>
> **H$_{1.3}$**   Webscraps helps users more effectively compare gathered information.

H$_2$ Participants find Webscraps more *engaging* than Page Thumbnails for information gathering tasks.

> **H$_{2.1}$**   Webscraps is more helpful with information gathering tasks.
>
> **H$_{2.2}$**   Users see added benefit in saving information as Webscraps versus Page Thumbnails.
>
> **H$_{2.3}$**   Webscraps is easier to learn than Page Thumbnails.
>
> **H$_{2.4}$**   Users enjoy using Webscraps more than Page Thumbnails.

5.2 USER STUDY

A user study was conducted during two weeks in July 2010. Thirty undergraduate and graduate students from Dalhousie Faculty of Computer Science (FCS) were recruited by email through a monitored "CSALL" mailing list (csall@cs.dal.ca) within FCS. To ensure a convenient and homogeneous sample, participants were required to be Computer Science students. Previous experience using the Mozilla Firefox Web browser was also mandated.

The study was conducted in the Usability Lab at FCS. Participants performed their tasks individually during a scheduled time and were not exposed to the other participants involved. All participants signed an Informed Consent form (Appendix A) and were notified prior to commencing that their $15 reimbursement was unconditional. There were no dropouts during the sessions.

5.3 STUDY DESIGN

This was a within-subject study in which participants used both Web browser gathering tools – *Webscraps* (WS) and *Page Thumbnails* (PT) – using two separate but similar gathering tasks of moderate difficulty: (1) "*Day Trip*" (Task A: Appendix B) and (2) "*Golf Trip*" (Task B: Appendix B). A summary of the study is given in Table 4.

Table 4: User Study Outline

Time (min.)	Procedure
5	Study Overview and Consent Form
5	Demographic Questionnaire
5	Warm-up Task for Tool 1
15	**Task 1**
5	Post-Task Questionnaire 1
5	Warm-up Task for Tool 2
15	**Task 2**
5	Post-Task Questionnaire 2
5	Post-Study Questionnaire
15	Post-study Semi-Structured Interviews

Tasks A and B were assigned in random order for each individual and required the participants to gather specific information from several websites over a period of 15 minutes. Prior to each task, an additional 5-minute training session demonstrated the features of

each tool while providing each user with a short warm-up task to reduce the bias of warm-up effects. Participants were told that information gathered may become useful in the future, and were therefore asked to perform information gathering activities as they would normally perform such tasks for themselves.

The interfaces were studied using a *within-subject* design (Table 5). The experimental design also evaluated the possible effects of *order* using four conditions in a *between-subject* design. Possible main effects of the independent variables were controlled by randomly assigning the order in which each interface (WS or PT) and task (A or B) was studied. Similar results were anticipated for opposite task order (A $\rightarrow$ B $\approx$ B $\rightarrow$ A) and interface order (PT $\rightarrow$ WS $\approx$ WS $\rightarrow$ PT).

Table 5: Study Design

Condition	Number of Participants	Description	
1	8	PT(A) $\rightarrow$ WS(B)	**Task**:
2	8	PT(B) $\rightarrow$ WS(A)	A = Day Trip B = Golf Trip
3	7	WS(A) $\rightarrow$ PT(B)	**Tool**:
4	7	WS(B) $\rightarrow$ PT(A)	WS = Webscraps PT = Page Thumbnails

The effects of independent variables (the tools, tasks and their order of study) were evaluated using a two-by-two-by-two (2 x 2 x 2) incomplete (partial) factorial design. Note that four conditions were studied rather than eight (2 x 2 x 2 = 8), as combinations in which participants would have repeated the same task or interface (e.g., PT(**A**) $\rightarrow$ WS(**A**)) were excluded.

Independent variables for each interface were measured using two criteria: *effectiveness* and *engagement*. Participants rated the effectiveness of each model during the second session by describing their degree of confidence in decision-making for the particular

task. Users filled-out a questionnaire in a 5-point Likert scale format with room for comments to provide an engagement rating.

At the beginning of the session, a background questionnaire was provided to gather participants' demographic information and background information using various Web browsing tools.

5.4 DATA COLLECTION

Four major data collection methods were used during this study: (1) a pre-study demographic questionnaire, (2) post-task and post-study questionnaires, (3) browser logs, and (4) post-study semi-structured interviews. All questionnaires were in Likert Scale format with room for additional written comments. Questionnaire and interview forms are available in Appendix A.

1. *Demographic Questionnaire*: A demographic paper questionnaire was administered at the beginning of the study. General demographic questions such as age, gender, experience using the Web and preferred Web browser were asked. Participants were reminded of their right to omit any information in the questionnaire which they felt uncomfortable answering.

2. *Post-Task and Post-Study Questionnaires*: Two identical post-task paper questionnaires were administered, one after each of the two tasks were completed. Questions pertaining to usability of both gathering tools and the assigned tasks were included. After completing both tasks and post-task questionnaires, participants were handed a post-study questionnaire with questions that compared both gathering tools and tasks.

3. *Browser Logs*: A desktop computer running Microsoft Windows provided by FCS was used to perform each task and collect Web browsing data. Two logging extensions for Mozilla Firefox were installed to monitor browsing behaviour:

"FoxyMeter"[3] monitored the total number of pages opened during each task, and "TabCounter"[4] logged the average and maximum number of tabs open per window.

A new Web browsing profile was created twice for each participant (one for each task) to ensure that Web history, Webscraps and Page Thumbnails data were kept separate, private and secure from other participants. Participants were asked to refrain from logging-in to any personal online accounts such as email during either of the sessions. Password-saving features within Firefox, and both the Adobe Flash and Adobe Portable Document Format (PDF) plug-ins were disabled.

4. *Post-study Semi-structured Interviews*: After both tasks and all questionnaires were completed, each participant was interviewed by means of three five-minute post-study semi-structured interviews to gather qualitative data. The first two interviews (Post-Task Semi-Structured Interviews) focused on one of the particular tasks-tool combination. The last interview (Post-Study Semi-Structured Interview) asked participants to contrast and compare both tools and tasks. Participants had access to the computer with their gathered Web data during each of the interviews.

---

[3] Available from: https://addons.mozilla.org/en-US/firefox/addon/3411/
[4] Available from: https://addons.mozilla.org/en-US/firefox/addon/6391/

CHAPTER 6   RESULTS

6.1 BACKGROUND QUESTIONNAIRE

Thirty participants from Dalhousie Computer Science and Health Informatics, 24 males and 6 females, between the ages of 19 and 37 (mean: 25.3, S.D.: 4.5) took part in the user study. Twenty-one participants were graduate students, while the remaining 9 were in undergraduate studies. All thirty participants reported using the Web more than five times a day for a combined average of 5.3 hours (S.D.: 1.29). Eighteen participants reported using the Web equally across all listed locations (school, home, and work; e.g., using a laptop), 5 participants reported using the Web primarily in two locations, while 3 used it at school, 3 at home only, and 1 at work only.

Participants were asked to indicate which browsers they had previously used (Table 6). Between 2 and 9 browsers were used in the past (mean: 4.93, S.D.: 1.53), while Mozilla Firefox, Microsoft Internet Explorer, and Google Chrome together constituted 57% of prior usage. Participants had been using their current primary browser for 2.96 years on average (SD: 2.55, min: "a few months", max: 10 years). These results indicate that the group had expertise using the Web and a variety of different web browsers.

Table 6: Web Browsers Previously Used

Participant	Web Browser							
	Mozilla Firefox	Microsoft IE	Google Chrome	Mozilla Classic	Apple Safari	Netscape Navigator	Opera	TOTAL
101	1	1	1	0	1	0	0	4
102	1	1	0	0	0	1	0	3
103	1	1	1	0	1	0	0	4
104	1	1	1	0	1	0	0	4
105	1	1	0	1	1	1	0	5
106	1	1	1	1	0	0	0	4
107	1	1	1	0	0	0	0	3

	Web Browser							
Participant	**Mozilla Firefox**	**Microsoft IE**	**Google Chrome**	**Mozilla Classic**	**Apple Safari**	**Netscape Navigator**	**Opera**	**TOTAL**
108	1	1	1	1	1	1	0	6
109	1	1	1	1	1	1	0	6
110	1	1	0	1	0	0	0	3
111	1	1	1	1	1	0	1	6
112	1	1	1	0	0	1	1	5
113	1	1	1	1	1	1	1	9*
114	1	1	1	1	1	0	1	6
115	1	1	1	0	1	1	1	6
116	1	1	1	0	1	0	1	5
117	1	1	1	1	1	1	1	7
118	1	1	1	1	0	1	1	6
119	1	1	1	1	0	0	0	4
120	1	1	1	1	1	0	0	5
121	1	1	1	1	0	1	1	6
122	1	1	1	0	1	1	0	5
123	1	1	1	1	1	1	1	7
124	1	1	0	0	0	0	0	2
125	1	1	0	0	0	0	0	2
126	1	1	1	0	0	0	1	4
127	1	1	1	1	0	1	0	5
128	1	1	1	1	0	0	1	5
129	1	1	1	1	1	0	0	5
130	1	1	1	1	0	1	1	6
Total	30	30	25	18	16	14	13	148
Mean	1.00	1.00	0.83	0.60	0.53	0.47	0.43	4.93
S.D.	0.00	0.00	0.38	0.50	0.51	0.51	0.50	1.38

A breakdown of the primary web browsers preferred by participants is shown in Figure 11. As of January 2011, the estimated market share of web browsers according to W3Counter[5] indicates Internet Explorer leading with a 56% share, Firefox 22.8%, Chrome 10.7%, and other web browsers including mobile platforms constituting the remaining 10.5%. The sample in this study preferred web browsers that diverged from the W3Counter results, particularly in their use of Internet Explorer. This may have been a result of the selection process (required experience using Mozilla Firefox) or a Computer

---

[5] "Global Web Stats", accessed Jan. 2011 <www.w3counter.com>

Science student bias toward cross-platform compatible browsers such as Firefox and Chrome.
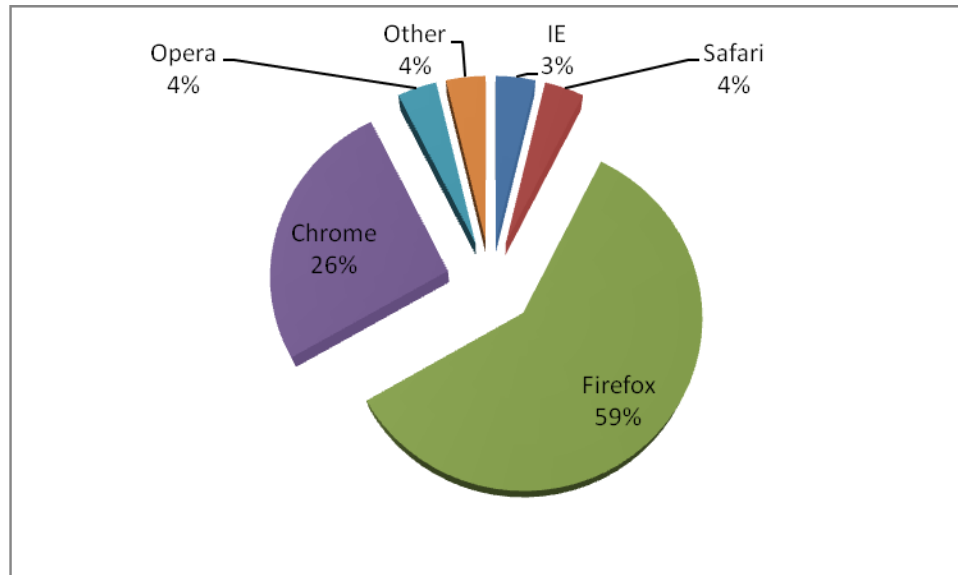
Figure 11: Primary Web Browser Used By Participants

Participants were asked how frequently they had performed three general categories of actions within a web browser: (1) Page Revisitation and Information Re-finding, (2) Information Gathering on the Web, and (3) Browser Tabs and Windows Use. The survey was administered using a 5-point Likert scale (1: "Never heard of" to 5: "Use Often"). Averages and standard deviations are presented in Table 7.

Table 7: Tools and Actions that Users Reported Within Web Browsers

Category	Tool/Action	Mean	S. D.
Page Revisitation and Information Re-finding	Use search engines	5.00	0
	Search within a web page	4.37	0.72
	Use Auto complete	4.37	0.89
	Use Bookmarks / Favorites	3.93	0.96
	Type a URL from memory	3.77	1.04
	Use History Lists	3.37	1.13
Information Gathering on the Web	Copy/Paste information to or from another application (e.g., Word, Notepad…)	4.17	0.83
	Email Information to myself or others	4.10	1.03
	Make notes about information I see (paper or electronic)	3.53	1.01

Category	Tool/Action	Mean	S. D.
	Save sessions	3.41	1.18
	Save web pages	2.97	1.13
	Print web pages	2.90	1.09
Use of Browser Tabs and Windows	Use tabs	4.93	0.37
	Open more than one browser window	4.20	0.81

Page Revisitation and Information Re-finding: In the first category, all participants reported using a search engine "often" (Mean: 5.00, S.D.: 0). Searching within a web page using the built-in browser search command (4.37, 0.72) and using the location bar or HTML form "auto-complete" function (4.37, 0.89) were the second most popular actions in the category. These three search-oriented strategies were more popular than using conventional bookmarks (3.93, 0.96), typing URLs from memory (3.77, 1.04), and using History lists (3.37, 1.13).

Information Gathering on the Web: For the second category, users preferred word processors and text editors (4.17, 0.83), and email clients (4.10, 1.03), to gather information online. Saving sessions (3.41, 1.18), saving web pages (2.97, 1.13), and printing web pages (2.90, 1.09) – strategies for saving entire web pages – were not rated above 4.

Use of Browser Tabs and Windows: Finally, participants used tabs often (4.93, 0.37) and opened more than one browser window during a session (4.20, 0.81).

The last two questions of the background questionnaire asked users to describe their overall information gathering activities. For the first question – "*Do you use the Web to gather information or special topics?*" – 29 participants specified that they used the Web to gather specific topics. Of these participants, 14 indicated an academic association (e.g., thesis literature search or programming projects), 2 participants associated information gathering with a social networking site (e.g., Facebook or YouTube), and 4 noted leisurely activities: reading the news, "tech-related stuff", "finding-out about movies" and "planning trips". For the second question – "*How would you start if you were planning a trip to Paris?*" – 17 participants reported that they would use a search engine such as

Google to perform an initial web search, while 8 participants preferred going directly to a travel search website (e.g., Expedia.com or Travelocity.com) by entering the URL manually. These results suggest a tendency to begin information gathering tasks by means of initial search queries.

6.2 POST-TASK QUESTIONNAIRES

Following each of the tasks, a post-task questionnaire containing 14 questions in 5-point Likert scale format was used to assess different characteristics of Webscraps, Page Thumbnails, and the tasks performed (Golf Trip and Day Trip). Table 8 reports the means and standard deviations of both tools for each of the 14 questions. Questions 5, 7, 8, 10, 11, and 12 requested additional explanations (e.g., "why or why not?", "for example", and "how?"). A qualitative categorization of these questions is listed in Appendix C.1.

Table 8: Post-task questionnaire results for Webscraps and Page Thumbnails

Question	Webscraps		Page Thumbnails	
	Mean	S.D.	Mean	S.D.
1. This tool was easy to use.	4.48	0.69	4.60	0.56
2. This tool was not hard to learn.	4.47	0.73	4.67	0.55
3. I would recommend this tool to other people.	3.87	0.94	3.57	0.90
4. I would like to see this tool as a standard feature in all Web browsers.	3.77	1.07	3.57	0.94
5. This tool contains useful new features.	**4.17**	**0.75**	**3.43**	**0.97**
6. This tool helped me complete my task.	3.77	0.94	3.79	0.86
7. I would use this tool for more challenging tasks.	3.80	1.06	3.37	1.30
8. I would use this tool for tasks that take longer to complete.	3.73	1.14	3.37	1.03

Question	Webscraps		Page Thumb- nails	
	Mean	S.D.	Mean	S.D.
9. This tool let me compare information from different websites.	**3.80**	**0.92**	**2.97**	**1.16**
10. I think this tool would help me remember information or websites.	4.03	0.93	3.57	1.10
11. This tool was useful for remembering important text.	**3.93**	**1.05**	**2.73**	**1.23**
12. This tool was useful for remembering important images.	4.00	0.86	3.53	1.11
13. I did not need more time to complete this task.	**3.03**	**1.45**	**3.83**	**1.21**
14. I have performed this type of task on the Web before.	2.80	1.24	3.07	1.34

The results of a paired samples t-test performed in SPSS between the Webscraps and Page Thumbnails post-task questionnaires demonstrated significant differences for questions 5, 9, 11 and 13 (Table 9). Participants significantly preferred Webscraps over Page Thumbnails for tasks that involved comparing information from different websites (question 9) and for remembering important text (q. 11). Interestingly, the ability to remember important images (q. 12) did not differ significantly between both tools.

Table 9: Significant Paired Samples t-Test Results

Question	t-Test Result
5. This tool contains useful new features.	$t = 4.43$; $df = 29$; $p < 0.000$
9. This tool let me compare information from different websites.	$t = 3.98$; $df = 29$; $p < 0.000$
11. This tool was useful for remembering important text.	$t = 5.07$; $df = 29$; $p < 0.000$
13. I did not need more time to complete this task.	$t = -3.525$; $df = 29$; $p < 0.001$

Participants perceived both tools as equally easy to *learn* (q. 2) and *use* (q. 1). These findings suggest that the additional selection step in Webscraps did not appear to affect usability. Both tools were highly rated in terms of their abilities to help users remember information or websites (q. 10). Both tools were also rated useful for completing the user study task (q. 6), more challenging tasks (q. 7) and tasks that would take longer to complete (q. 8).

Interestingly, question 13 revealed an opposite trend. Participants required on average more time to complete their tasks using Webscraps than Page Thumbnails. It is suspected that the extra selection step using Webscraps meant that participants spent more time deciding what specific information was useful within pages, thereby increasing the overall time necessary to complete the task. A subsequent study would be necessary to investigate differences in *efficiency* between these tools.

6.3 POST-SESSION SURVEY

The post-session survey contained 10 questions that directly compared Webscraps to Page Thumbnails using the same 5-point Likert scale format used for the post-task questionnaire. To facilitate interpretation of the results, the scale was collapsed into three categories: "agree", "indifferent" and "disagree". Likert answers 1 and 2 ("strongly disagree" and "disagree") were considered as general disagreements to a question, whereas 4 and 5 ("agree" and "strongly agree") were considered as general agreements, and 3 was regarded as an indifferent opinion. These were used to accommodate to the different valuations between users. The results from the post-session survey are listed in Table 10. Five questions (3, 4, 8, 9 and 10) in the post-session survey requested additional comments. A qualitative categorization of participant comments is summarized in Appendix C.2.

Table 10: Post-Session Survey Results (Agreements ≥ 70% shown in bold)

Questionnaire Questions	Agree	Indifferent	Disagree
	% (#)	% (#)	% (#)
1. Page Thumbnails was not easier to use than Web Scraps.	20% (6)	23% (7)	53% (16)
2. I enjoyed using Web Scraps more than Page Thumbnails.	57% (17)	23% (7)	17% (5)
3. Page Thumbnails was not more helpful for complex tasks than Web Scraps.	60% (18)	17% (5)	17% (5)
4. Page Thumbnails was not easier to learn than Web Scraps.	17% (5)	33% (10)	47% (14)
5. Web Scraps helped me complete my task more efficiently than Page Thumbnails.	63% (19)	17% (5)	17% (5)
6. I do not think Page Thumbnails helped me remember more relevant information than Web Scraps.	60% (18)	20% (6)	17% (5)
7. Web Scraps helped me compare information between websites more easily than Page Thumbnails	**70% (21)**	**13% (4)**	**13% (4)**
8. I do not think I can recall specific information easier by looking at Page Thumbnails than Web Scraps.	**73% (22)**	**10% (3)**	**13% (4)**
9. I do not think I can recall entire websites easier using Page Thumbnails than Web Scraps.	20% (6)	20% (6)	57% (17)
10. I think Web Scraps helped me remember why I chose to keep something more efficiently than Page Thumbnails.	**73% (22)**	**20% (6)**	**1% (1)**

The post-session survey results indicate that users preferred Webscraps over Page Thumbnails for comparing information between websites (q. 7, 21/30 = 70%), recalling specific information (q. 8, 22/30 = 73%) and helping to remember why information was saved (q. 10, 22/30 = 73%). While 53% (16/30) of people agreed that Page Thumbnails was easier to use (q.1) and 47% (14/30) thought it was easier to learn (q. 4), both questions had comparably high indifference scores (q. 1, 23% = 7/30; q. 4, 33% = 10/30),

suggesting that both tools were reasonably easy to learn and use. These statements are consistent with earlier results (Questions 1 and 2) from the post-task questionnaires.

Sixty percent (18/30) of participants agreed that Webscraps was more helpful than Page Thumbnails for complex tasks (q. 3). Within those that agreed, 66% (12/30) of participants noted in the written portion of Question 3 that Webscraps' ability to focus on a specific area of a website was a feature that assisted with complex tasks. A similar percentage (63% = 19/30) agreed that Webscraps helped complete the task more efficiently than Page Thumbnails (q. 5).

6.4 WEB BROWSER DATA

Two additional third-party Firefox browser extensions were used during the sessions to track browsing patterns. TabCounter[6] tracked the highest number of tabs opened simultaneously per session and FoxyMeter[7] counted the total number of web pages visited (Table 11).

Table 11: Web Browser Data

	Highest number of tabs open simultaneously		Total number of pages visited	
	WS	PT	WS	PT
Min	2	3	51	47
Median	6	7	82.5	74.5
Max	16	16	164	126
Mean (S.D.)	7.10 (3.71)	7.83 (3.23)	87.50 (28.00)	78.17 (21.58)

On average, participants visited more pages during their sessions with Webscraps than with Page Thumbnails. A paired-samples t-test revealed this finding to be a significant

---

6 Source: <https://addons.mozilla.org/en-US/firefox/addon/6391/>
7 Source: <https://addons.mozilla.org/en-US/firefox/addon/3411/>

difference ($t = 2.10$; $df = 29$; $p < 0.05$). These results follow earlier findings (Question 13 in the Post-Task Questionnaire) where participants using Webscraps spent significantly more time completing tasks.

Another t-test comparing the maximum number of simultaneous tabs per tool revealed no significant differences between Webscraps and Page Thumbnails. Possible significant differences between Task A and Task B (Day Trip and Golf Trip) were also tested for in further t-tests. No significant differences were found for page and tab counts between Task A and Task B.

CHAPTER 7   DISCUSSION

This section discusses data from the user study in order to accept or reject the hypotheses and sub-hypotheses introduced in the Methodology chapter. Questionnaires and interviews are available for reference in Appendix A.

H.1: WEBSCRAPS IS A MORE EFFECTIVE GATHERING TOOL THAN PAGE THUMBNAILS FOR INFORMATION GATHERING TASKS.

H.1.1: Webscraps increases users' ability to remember gathered information.

From an analysis of questionnaire and interview results, Sub-Hypothesis H.1.1 was found to be true.

Text: Based on a paired samples t-test of Likert scale data, Webscraps was *significantly* preferred (t = 5.07; df = 29; p < 0.000) over Page Thumbnails for remembering important text (q. 2-11).

In post-task questionnaires, 50% (15/30) of participants using Webscraps noted an ability to "focus on", "select" or "make snapshots" of specific areas of text. Among these results, participants wrote "specific information [was] easy to save", "make a snapshot of the page and it shows where you highlighted", and "[Webscraps] could store the text I was interested in". One participant after using Webscraps even noted "I think [Webscraps] could be useful for keeping lists of things". Later this was supported in the post-task interview, when a participant noted selecting visual "headings of pages where the gathered information was found" instead of the content of interest.

Twenty-three percent (7/30) of participants noted that Page Thumbnails was not useful for remembering text, while another 23% (7/30) of participants commented that web page text in Page Thumbnails was too small to read. One comment read "[I] would have to re-

open the tab to read [the web page]", while another stated "[Page Thumbnails] doesn't allow you to choose important parts in a website, other than the title of the website, there's not much information or 'text'". These comments reflect on the difference between Webscraps and Page Thumbnails: Webscraps selections are not reduced by scaling but maintain their original size and resolution within their interface; Page Thumbnails, however, are 200 x 400 pixel (height x width) thumbnails of a reduced web browser's view-port.

Images: Based on the paired samples t-test, neither tool was significantly preferred for remembering important images.

In the question commentary, 37% (11/30) of participants thought that Webscraps was useful for images. One participant noted "select only important images", while another wrote that selection was the "same as for text". Thirteen percent (4/30) noted that they had gathered few or no images. A participant using Webscraps wrote "most of my windows ended up being text", while another noted, for images "I think [Webscraps] would be [useful], but I didn't really have any important images". Another participant wrote: "[m]ostly getting screen shot[s] of important text, however [Webscraps] was handy for maps". During the tasks, it was observed that Google Maps was searched on more than one occasion to select a region of interest.

Participants using Page Thumbnails had similar comments. Thirty-three percent (10/30) found pictures easily viewable within Page Thumbnails. One user wrote "Images on web pages are larger than the text... they are easier to remember". Twenty-three percent (7/30) of participants gathered few or no images. One user noted Page Thumbnails were "[m]ore so for saving an entire page than a particular image", while two participants wrote "no relevant images for this task".

Recall and Relevant Information: In terms of *recalling* specific information, 73% (22/30) of participants preferred Webscraps in post-task surveys, while 13% (4/30) preferred Page Thumbnails and 10% (3/30) were undecided. Sixty percent (18/30) of par-

ticipants preferred Webscraps over Page Thumbnails (17%, 5/30) for remembering more *relevant* information, while 20% (6/30) of users were "indifferent".

In the commentary, 40% (12/30) of participants wrote that highlighting important areas of a page made Webscraps a more effective tool for recalling specific information. Three participants found it easier to remember what they had selected. For example, "[i]n [Page Thumbnails] I have to remember why I went to that site in the first place. In Webscraps I could bookmark [the] area of interest." For remembering *why* they had chosen information, 73% (22/30) of participants preferred Webscraps, 3% (1/30) Page Thumbnails, and 20% (6/30) were undecided. Fifty-three percent (16/30) of users commented that Webscraps highlights or focuses information, making it "constantly in view", or "[y]ou see exactly what it was that made you want to save it, instead of the entire site."

Remembering Information or Websites: No significant preference was found for either tool for remembering "information or websites".

A lack of significant preference toward a particular tool for remembering information or websites is consistent other results in this study: (1) Webscraps were significantly preferred for gathering specific information such as text, while (2) Page Thumbnails were preferred for recognizing websites. As such, responses to this question were mixed – there was no significant preference toward a single tool to do both tasks. One participant summarized this finding in his/her comments by noting: "[Webscraps] focuses in on the specific piece of information I want to remember, not just the whole webpage". Later, while using Page Thumbnails, the same participant wrote "[Page Thumbnails] [w]ould help me remember websites better – but not information, as tool allows me to get repeated sightings of the layout / colour of the webpage, while the text is too small to see".

Summary H.1.1: Webscraps increases users' ability to remember gathered information.

- ***Text***: Webscraps significantly increases users' ability to remember gathered information in the form of text.

- ***Images***: No significant preference toward either tool for remembering images.

- ***Recall and Relevant Information***: Webscraps was preferred for recalling specific information (73% = 22/30) and for remembering relevant information (60% = 18/30).

- ***Remembering Information or Websites***: Neither tool was significantly preferred for remembering *both* specific information and recalling websites.

Recommendations H.1.1:

- During information gathering tasks, providing a means to *gather specific information* from within web pages will help users more effectively remember gathered information.

- Scaled text (content) is less recognisable than images scaled to the same degree.

- Tools that capture images and document headings provide visual cues that aid in recognizing information that has been gathered.

- Users find it useful to keep information gathered online within the web browser.

H.1.2: Webscraps helps users more effectively recognize websites.

From an analysis of questionnaire and interview results, Sub-Hypothesis H.1.2 was found to be false.

Recognizing Websites: In post-task surveys, 57% (17/30) of participants preferred Page Thumbnails for recognizing websites. Twenty percent (6/30) preferred Webscraps and 20% (6/30) were indifferent (1 participant did not answer).

Thirty-three percent (10/30) of participants commented that seeing an entire web page in Page Thumbnails allowed for better page recognition than just seeing a selection in Web-

scraps. Immediately following the tasks, one participant wrote "[Page] Thumbnails gives a general view of the web page, so you can see the entire layout", while another user wrote: "I don't think [Page Thumbnails] was useful for remembering 'important' text but [it] helped [me] remember what pages [were] about". Interestingly, two participants found that information unrelated to the gathering task provided visual cues to aid in web page recognition: "irrelevant [web page] data can give me more recall for relevant [web page data]".

Layers in Webscraps: Ten percent (3/30) of participants mentioned benefiting from the Webscraps "Layer" background at some point during tasks. One participant noted that Webscraps helped remember information or websites "[b]y saving the information in the window as well as in the background", while another wrote "Web Scraps shows the entire page in the background". During interviews, another participant wrote "[I] compared table vs. Category view (golf balls); for box 6 I opened up the background to see more prices". Others, when asked about Layers noted "Webscraps had a few non-obvious features" and "not all the useful features [of Webscraps] were used in the first task". Nonetheless, two participants showed interest in using Layers: "[I] liked 'Layer Peek' but [I] didn't use it during the short time" and "the background was useful; for short tasks I would not really use the background; Webscraps is better for longer tasks". Two participants disliked the layer, noting that "the background was distracting" and "the background [in Webscraps] should be white" as in the case of Page Thumbnails. At this stage, use of the Webscraps Layer for recognizing web pages requires additional study.

Summary H.1.2: Webscraps helps users more effectively recognize websites.
- ***Recognizing Websites***: Both quantitative and qualitative data indicate that users found Page Thumbnails more effective than Webscraps for recognizing web pages.
- ***Layers in Webscraps***: Further study with greater emphasis on Webscraps Layer's evaluation is necessary to establish its possible value for helping recognize web pages.

Recommendations H.1.2:

- To help with page recognition, participants prefer seeing the entire layout (initial browser viewport) of a web page as opposed to selections of information (e.g., Webscrap).

- Provide a means to zoom-in and perform textual search within thumbnails of web pages.

H.1.3: Webscraps helps users more effectively compare gathered information.

From an analysis of questionnaire and interview results, Sub-Hypothesis H.1.3 was found to be true.

Comparing information: A paired samples t-test of Likert scale data revealed that participants *significantly* preferred Webscraps over Page Thumbnails ($t = 3.98$; $df = 29$; $p < 0.000$) for comparing information from different websites. Seventy percent (21/30) of participants agreed that Webscraps helped them compare information more easily than Page Thumbnails, while 13% (4/30) preferred Page Thumbnails and the remaining 13% (4/30) were indifferent.

Twenty percent (6/30) of participants using Webscraps and 17% (5/30) using Page Thumbnails reported comparing and combining as useful new features. Among them, one participant wrote "[I] like how everything [in Webscraps] is in one place", another Webscraps user wrote "Specific areas are selected... information is together for comparison". A participant using Page Thumbnails wrote "Quick view comparison feature... didn't need to go back to pages", while another Page Thumbnail user wrote "Small format of entire trip on one page (e.g., bus, food...)".

Different Methods of Comparison: Three participants preferred using browser tabs instead of Webscraps and Page Thumbnails for comparison (WS: 1, PT: 2). Users noted that "[i]t is simpler to compare pages using tabs [than using Webscraps]" and "[Page

Thumbnails] doesn't feel much different than just tabbing the pages out in the browser".
Two other participants wrote that Webscraps was "…a nice alternative to leaving all
kinds of tabs open – just save the bits you need and it's all in one tab" and "for making
comparisons… it's nice to have this information in a web browser". While the effective-
ness of comparison between Webscraps and Page Thumbnails was studied, effectiveness
of comparison between Webscraps and browser tabs (and Page Thumbnails and browser
tabs) requires additional study.

Summary H.1.3: Webscraps helps users more effectively compare gathered information.

- *Comparing information*: Participants *significantly* preferred Webscraps over
  Page Thumbnails for comparing information.

- *Different Methods of Comparison*: Effectiveness of comparison between Web-
  scraps and browser tabs (and Page Thumbnails and tabs) requires additional
  study.

Recommendations H.1.3:

- Users enjoy the ability to compare selections of data within a web browser.

- Provide a means to rearrange information within an interface after it has been
  gathered.

- Provide the ability to zoom in on selections in cases thumbnail scaling is used.

H.2: PARTICIPANTS FIND WEBSCRAPS MORE ENGAGING THAN PAGE THUMB-NAILS FOR INFORMATION GATHERING TASKS.

H.2.1: Webscraps is more helpful with information gathering tasks.

From an analysis of questionnaire and interview results, Sub-Hypothesis H.2.1 was found
to be inconclusive.

Information Gathering Tasks: A paired samples t-test of post-task questionnaires re-
vealed no significant differences between Webscraps and Page Thumbnails for helping to
complete tasks. Interestingly, 63% (19/30) of participants agreed that Webscraps helped

them complete tasks more *efficiently* than Page Thumbnails (17% = 5/30; the remaining 17% were 'indifferent'). Sixty percent (18/30) of participants also agreed that Webscraps was more helpful for *complex tasks* (3-3) than Page Thumbnails (17% = 5/30; 17% were 'indifferent'). Paired samples t-tests were run on post-task questionnaire data to determine whether users significantly preferred either tool for more *challenging* tasks or tasks that take *longer to complete*. Results from these two tests found no significant preference for either tool.

Complex Tasks (commentary): Forty-seven percent (14/30) of participants noted that Webscraps helped with complex tasks by means of helping to focus on specific information or important areas within web pages. Among these responses, one participant wrote "The [Webscrap] really narrows down on the information. I think [as] implemented, Webscraps are for projects, [T]humbnails for bookmarks". One participant revealed that making selections using Webscraps helped them think about the information they were selecting for their task: "Webscraps was cool in psychological effect involving user to look for the spot he/she wants". Ten percent (3/30) of participants noted that while using Webscraps, original web pages did not need to be revisited to see gathered information. Two participants wrote that Page Thumbnails' simplicity would be more helpful in complex tasks.

Challenging Tasks (commentary): For more challenging tasks, 47% (14/30) of participants noted Webscraps was more useful than Page Thumbnails (23% = 7/30). Reasons for using Webscraps included saving specific information (13% = 4/30), helping to remember why information was saved (13% = 4/30), and the ability to perform more effective comparisons (10% = 3/30). Among the reasons for using Page Thumbnails, 17% (5/30) noted the ability to save and remember information. Twenty percent (6/30) of Webscraps users and 30% (9/30) of Page Thumbnails users found that the tools were not useful for more complex tasks, while 13% (4/30) and 20% (6/30) were undecided of using Webscraps and Page Thumbnails for complex tasks. Reasons for not using Webscraps were varied (e.g., "I think text-based lists are better... [Webscraps] could be useful for something more visual"; "Didn't seem much more useful than just switching tabs.").

For Page Thumbnails, three participants preferred using traditional bookmarks for more challenging tasks.

Summary H.2.1: Webscraps is more helpful with information gathering tasks.

- ***Information Gathering Tasks***: No significant differences (paired sample t-tests) were found between Webscraps and Page Thumbnails for helping to complete (1) normal tasks, (2) challenging tasks, or (3) tasks that take longer to complete.
- ***Complex Tasks (Commentary)***: Forty-seven percent (14/30) of participants preferred Webscraps for its ability to focus on specific information or important areas within web pages.
- ***Challenging Tasks (Commentary)***: Qualitative responses indicate a preference toward using Webscraps for more challenging tasks.

Recommendations H.2.1:

- The ability to gather *specific* information helps reduce clutter during complex information gathering tasks.

H.2.2: Users see added benefit in saving information as Webscraps versus Page Thumbnails.

From an analysis of questionnaire and interview results, Sub-Hypothesis H.2.2 was found to be true.

Useful New Features: A paired samples t-test revealed that participants *significantly* preferred ($t = 4.43$; $df = 29$; $p < 0.000$) Webscraps over Page Thumbnails in terms of useful new features.

Webscraps' Features: Forty percent (12/30) of participants using Webscraps noted that selecting specific information or areas within pages was a useful new feature. One participant wrote "[Webscraps] help[s] the users to bookmark a particular information on a search web page for future use", while another user noted, "I only ever bookmark pages

for a specific reason. Something like this would make long lost bookmarks a thing of the past." Seventeen percent (5/30) of users enjoyed the efficiency of Webscraps. For example, one participant noted, "[p]inpoint exact price/phrase and com[e] back directly to it"; another user wrote, "It is an interesting tool. Not only saves my web pages but I can look at info I want in a glance without opening up a page". Thirteen percent (4/30) of users noted Webscraps' ability to combine related information from several sources into one place. One participant noted, "[y]ou can put information in close proximity to other relevant information", while two other users wrote, "[Webscraps] [h]elps to condense a complex session into a single tab for organization and review" and, "[y]ou do not need to copy and paste to a Word document anymore".

Page Thumbnails' Features: Twenty percent (6/30) of participants wrote that Page Thumbnails provided a more visual experience than using browser windows, tabs or bookmarks. For example, a participant using Page Thumbnails wrote, "[m]ore visual than tabs/mutiple windows being open", while another comment read, "[s]eem[ed] useful to be able to bookmark with images". One user wrote, "[Page Thumbnails] has useful features including search result summarization by means of pictures". Thirty percent (9/30) of participants had previously used or seen a browser tool similar to Page Thumbnails, while none had reported seeing a tool similar to Webscraps. Among the results, 3 participants compared Page Thumbnails to Google Chrome's "Thumbnails", 3 reported a similarity to Opera's "Speed Dial", and 2 compared Page Thumbnails to existing Firefox extensions. Twenty-three percent (7/30) of participants using Page Thumbnails did not find the tool useful, compared to a single participant using Webscraps.

Useful in which situations? Thirty percent (9/30) of participants noted that, depending on the situation, both Webscraps and Page Thumbnails were useful. Thirty-three percent (10/30) of participants preferred Webscraps in situations requiring work with specific information. For example, two participants noted gathering information during academic tasks such as a "literature review" or "writing a paper". Thirteen percent (4/30) of participants noted Webscraps' use in comparing product prices. These comments illustrate Webscraps' usefulness in information gathering sessions where a user is looking to gather

specific information. The user may or may not have a well-defined goal, however, they are already in the process of analyzing specific information.

In the case of Page Thumbnails, 23% (7/30) responses shared a common theme of *page revisitation*. For example, 2 participants noted Page Thumbnails was useful in gathering "simpler information" or during "simple tasks". One user noted that Page Thumbnails would be useful in "more general tasks", while another noted its use for "overview of searches", and still another its ability to "save quickly for later to eventually get to, no specifics" [sic]. These comments illustrate the use of Page Thumbnails in information gathering situations where the specific goal of the task may or may not be well established, however, the user wants to gather pages for a more in-depth analysis at a later date. Two participants noted that Page Thumbnails was useful for pages that are "frequent revisits". For example, one user commented: "[Webscraps is] good for static pages with specific information; [Page Thumbnails for]: dynamic pages (e.g,: Gmail)".

Summary H.2.2: Users see added benefit in saving information as Webscraps versus Page Thumbnails.

- ***Useful New Features***: Participants *significantly* preferred Webscraps over Page Thumbnails in terms of useful new features.
- ***Webscraps' Features***: Forty percent (12/30) of participants using Webscraps noted that selecting specific information or areas within pages was a useful new feature.
- ***Page Thumbnails' Features***: Twenty percent (6/30) of participants wrote that Page Thumbnails provided a more visual experience than using browser windows, tabs or bookmarks.
- ***Useful in which situations?*** Thirty-three percent (10/30) of participants preferred Webscraps in situations that required working with *specific information*. In the case of Page Thumbnails, 23% (7/30) of participants valued Page Thumbnails during *page revisitation*.

Recommendations H.2.2:

- For tasks in which users gather and work with specific information, tools should provide the ability to select information from within pages as well as preview what has been selected. Previewing helps with comparison and recall of specific information.

- For more *general* information gathering tasks (no specific information being gathered) or for saving frequently-revisited pages, web page thumbnails are better suited than thumbnails of selections.

H.2.3: Webscraps is easier to learn than Page Thumbnails.

From an analysis of questionnaire and interview results, Sub-Hypothesis H.2.3 was found to be inconclusive, as both tools were easy to learn and use.

Ease of Use and Ease of Learning: Results of paired samples t-tests on post-task questionnaires revealed no significant differences between Webscraps and Page Thumbnails in terms of ease of use and ease of learning. Fifty-three percent (16/30) of participants agreed that Page Thumbnails was easier to use, while 20 % (6/30) disagreed and 23% (7/30) were indifferent. Forty-seven percent (14/30) of participants agreed that Page Thumbnails was easier to learn than Webscraps, while 17% (5/30) disagreed and 33% (10/30) were indifferent.

Twenty-seven percent (8/30) of participants commented that both tools were of similar difficulty, while 10% (3/30) noted that both tools were easy to learn. While 23% (7/30) of the participants noted that Webscraps contained more features than Page Thumbnails (or Page Thumbnails contained fewer features than Webscraps), 3 of the responses agreed that both tools were either "both easy to use" or of "similar difficulty". Seventeen percent (5/30) of participants wrote that Page Thumbnails required fewer steps to gather information (e.g., "[Webscraps] has extra step of area capture").

Summary H.2.3:

- *Ease of Use and Ease of Learning*: No significant differences in terms of ease of use and ease of learning were found between Webscraps and Page Thumbnails.

- 27% (8/30) of participants commented that both tools were of similar difficulty.

H.2.4: Users enjoy using Webscraps more than Page Thumbnails.

From an analysis of questionnaire and interview results, Sub-Hypothesis H.2.4 was found to be true.

Enjoyment: Fifty-seven percent (17/30) of participants enjoyed using Webscraps more than Page Thumbnails (17% = 5/30), while 23% (7/30) were indifferent. A paired samples t-test of post-task questionnaires revealed no significant preference between either tool for the following categories: "I would recommend this tool to other people" and, "I would like to see this tool as a standard feature in all web browsers".

Sixty percent (18/30) of participants commented that they enjoyed using Webscraps more than Page Thumbnails, while 23% (7/30) preferred Page Thumbnails (5-2). Four participants commented that Webscraps was more enjoyable to use because it provided a means to focus or concentrate on specific information. For example, one participant noted, "[I am] able to concentrate on the important part of my search. Coming back after a few days, I would not [otherwise] be aware of where the important information is located." Three participants preferred Page Thumbnails due to a simpler interface ("less effort required", "fewer distractions" and "more user friendly").

Summary H.2.4:

- *Enjoyment*: Fifty-seven percent (17/30) of participants enjoyed using Webscraps more than Page Thumbnails, while 17% (5/30) preferred Page Thumbnails (Likert data).

- From the commentary, 60% (18/30) of participants enjoyed using Webscraps more than Page Thumbnails, while 23% (7/30) preferred Page Thumbnails.

A summary of the most important comments from all questionnaires and interviews (Table 12) revealed interesting trends between Webscraps and Page Thumbnails (for the full summary of comments see Appendix C.5). Overall, there were 99 more comments relating to Webscraps than Page Thumbnails (311 and 212 comments, respectfully). The most recurring comment for Webscraps – *Selecting specific information/areas* (41% = 127/311) – was also three times more frequent (or twice as frequent percentage-wise) than the most recurring comment for Page Thumbnails – *Remembering websites* (18% = 39/212).

Table 12: A Comparison of the Most Frequent Comments

Rank	Webscraps Comments	Count	Page Thumbnails Comments	Count
1	**Selecting specific information/areas**	**127**	**Remembering websites**	**39**
2	**Comparing information**	**26**	Similar to another tool / I have used this tool before	16
	…			
9	Layers (Background) was helpful	7	**Comparing information**	**7**
	…			
16	**Remembering websites**	**5**	**Saving/Remembering information**	**6**
	…			
	TOTAL WS COMMENTS	**311**	**TOTAL PT COMMENTS**	**212**

Comparison of Most Frequent Comments: Overall, participants found that selecting specific information or areas within Webscraps was by far the most important feature of both tools. The corresponding comment from Page Thumbnails – *Saving/Remembering information* – was negligible by comparison (3% = 6/212). *Comparing information*, was three times more frequent in Webscraps (8% = 26/311) than Page Thumbnails (3% = 7/212). Page Thumbnails, which excelled at *Remembering websites*, contained eight times more corresponding comments (18% = 39/212) than Webscraps (2% = 5/311). These qualitative results for (1) *Selecting specific information/areas*, (2) *Comparing information*, and (3) *Remembering websites* are consistent with corresponding significant results obtained from post-task questionnaires.   This agreement also assures that the qualitative data obtained from commentaries and interviews follow the same trends as quantitative data obtained from Likert scale questionnaires.

DISCUSSION SUMMARY

A summary of the hypotheses and sub-hypotheses is given by Table 13.

Table 13: Summary of Hypotheses and Sub-Hypotheses

Hypothesis or Sub-hypothesis		Result
H$_1$	**Webscraps is a more effective gathering tool than Page Thumbnails.**	✔
H$_{1.1}$	Webscraps increases users' ability to remember gathered information.	✔
H$_{1.2}$	Webscraps helps users more effectively recognize websites.	✘
H$_{1.3}$	Webscraps helps users more effectively compare gathered information.	✔
H$_2$	**Participants find Webscraps more engaging than Page Thumbnails**	✔
H$_{2.1}$	Webscraps is more helpful with information gathering tasks.	?
H$_{2.2}$	Users see added benefit in saving information as Webscraps versus Page Thumbnails.	✔
H$_{2.3}$	Webscraps is easier to learn than Page Thumbnails.	?
H$_{2.4}$	Users enjoy using Webscraps more than Page Thumbnails.	✔

Guidelines for Information Gathering Tools

For developers wishing to implement a tool to perform Page Revision, Information Re-finding, and Information Combination, the following guidelines are provided:

For Page Revisitation

- Provide help with page recognition by preserving the entire layout of a page (not just selections of a page) within a thumbnail.

- Provide visual cues (e.g., page headings and images) to help recognize information.

- To assist with readability of thumbnails of pages, provide pan and zoom capabilities.

For Information Re-finding

- Provide means to help users gather and save only relevant information within a web page (e.g., selections of pages).

For Information Combination

- Provide means to compare information (e.g., tables) between pages.

CHAPTER 8   CONCLUSION

8.1 OVERVIEW AND LIMITATIONS

This thesis presented Webscraps, a tool to assist with three subtasks of information gathering: *information re-finding*, *page revisitation*, and *information combining*. Webscraps was designed based on guidelines proposed in background research and implemented as a Mozilla Firefox web browser extension.

A lab study evaluated Webscraps against Page Thumbnails, a clone of visual bookmarking interfaces increasingly popular in most major web browsers, interfaces which unlike Webscraps do not allow for selection of specific information within a web page. A sample size of 30 was chosen to provide a minimum of 7-8 participants for each of four conditions used. Counterbalancing was performed and each participant was randomly assigned to a condition. Two main hypotheses relating to effectiveness and engagement were tested along with seven additional sub-hypotheses.

A convenient sample of Computer Science students from Dalhousie University was selected for the user study. Although not representative of the general web browsing population, CS students require less technical training prior to tasks and are capable of providing more detailed feedback. While tasks were not self-motivated, semi-structured interviews allowed participants to express any additional comments or suggestions with researchers.

8.2 RESULTS

User study results based on paired samples t-tests indicate that participants *significantly* preferred Webscraps over Page Thumbnails for information gathering tasks that involve *comparing information from different websites* and *remembering important text*. Of the

two hypotheses and seven sub-hypotheses outlined in the Methodology, both main hypotheses were accepted – *Webscraps is a more effective gathering tool than Page Thumbnails* and *Participants find Webscraps more engaging than Page Thumbnails*. Among the sub-hypotheses, participants found that Webscraps increased their ability to *remember gathered information* and *effectively compare gathered information*. Page Thumbnails were more effective at *recognizing web pages*. For information gathering tasks, both tools were found to have their own merits, as so neither tool was found more helpful with *information gathering tasks* in general. Finally, both tools were of similar difficulty in terms of a learning curve.

8.2 FUTURE WORK

Future work will focus on other information gathering subtasks (e.g., decision-making) related to Webscraps' potential use in multi-session web browsing. As well, additional studies will evaluate Webscraps with users and their own tasks in a field evaluation.

BIBLIOGRAPHY

Abrams, D., & Baecker, R. (1997). How people use WWW bookmarks. *CHI '97: CHI '97 extended abstracts on Human factors in computing systems* (pp. 341-342). New York, NY, USA: ACM. doi: http://doi.acm.org/10.1145/1120212.1120424.

Abrams, D., Baecker, R., & Chignell, M. (1998). Information archiving with bookmarks: personal Web space construction and organization. *CHI '98: Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 41-48). New York, NY, USA: ACM Press/Addison-Wesley Publishing Co. doi: http://doi.acm.org/10.1145/274644.274651.

Adar, E., Teevan, J., & Dumais, S. T. (2008). Large scale analysis of web revisitation patterns. *CHI '08: Proceeding of the twenty-sixth annual SIGCHI conference on Human factors in computing systems* (pp. 1197-1206). New York, NY, USA: ACM. doi: http://doi.acm.org/10.1145/1357054.1357241.

Adar, E., Teevan, J., & Dumais, S. T. (2009). Resonance on the web: web dynamics and revisitation patterns. *CHI '09: Proceedings of the 27th international conference on Human factors in computing systems* (pp. 1381-1390). New York, NY, USA: ACM. doi: http://doi.acm.org/10.1145/1518701.1518909.

Alexander, J., Cockburn, A., Fitchett, S., Gutwin, C., & Greenberg, S. (2009). Revisiting read wear: analysis, design, and evaluation of a footprints scrollbar. *CHI '09: Proceedings of the 27th international conference on Human factors in computing systems* (pp. 1665-1674). New York, NY, USA: ACM. doi: http://doi.acm.org/10.1145/1518701.1518957.

Amin, A. (2009). Establishing requirements for information gathering tasks. *CWI. Information Systems [INS]*, 1–8. Citeseer. Retrieved January 13, 2011, from http://scholar.google.com/scholar?hl=en&btnG=Search&q=intitle:Establishing+requirements+for+information+gathering+tasks#0.

Aula, A., Jhaveri, N., & Käki, M. (2005). Information search and re-access strategies of experienced web users. *WWW '05: Proceedings of the 14th international conference on World Wide Web* (pp. 583-592). New York, NY, USA: ACM. doi: http://doi.acm.org/10.1145/1060745.1060831.

Boardman, R., & Sasse, M. A. (2004). "Stuff goes into the computer and doesn't come out": a cross-tool study of personal information management. *CHI '04: Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 583-590). New York, NY, USA: ACM. doi: http://doi.acm.org/10.1145/985692.985766.

Bruce, Harry, Jones, William, & Dumais, Susan. (2004). Keeping and re-finding information on the web: What do people do and what do they need? *Proceedings of the American Society for Information Science and Technology*, *41*(1), 129-137. doi: 10.1002/meet.1450410115.

Choo, C. W., Detlor, B., & Turnbull, D. (2000). Information Seeking on the Web: An Integrated Model of Browsing and Searching. *First Monday*, *5*(2).

Cockburn, A., Greenberg, S., Jones, S., Mckenzie, B., & Moyle, M. (2003). Improving Web Page Revisitation: Analysis, Design, and Evaluation. *IT & Society*, *1*, 159-183. doi: 10.1.1.19.9613.

Ellis, D. (1993). Modeling the Information-Seeking Patterns of Academic Researchers: A Grounded Theory Approach. *The Library Quarterly*, *63*(4), 469-486. Retrieved from http://www.jstor.org/stable/4308867.

Google/Inc. (n.d.). Google Chrome Web Browser. Retrieved February 2, 2010, from http://www.google.com/chrome/.

Hiner, K. (2009). Safari 4 Top Sites first look. *http://www.appletell.com*. Retrieved February 2, 2010, from http://www.appletell.com.

Jones, W, Bruce, H, & Dumais, S. (2003). How do people get back to information on the web? How can they do it better. *9th IFIP TC13 International Conference on Human-Computer Interaction (INTERACT 2003), Zurich, Switzerland*.

Jones, William, Bruce, Harry, & Dumais, Susan. (2001). Keeping found things found on the web. *CIKM '01: Proceedings of the tenth international conference on Information and knowledge management* (p. 119--126). New York, NY, USA: ACM. doi: http://doi.acm.org/10.1145/502585.502607.

Kellar, M., Watters, C., & Shepherd, M. (2007). A field study characterizing Web-based information-seeking tasks. *Journal of the American Society for Information Science and Technology*, *58*(7), 999-1018. Retrieved from http://portal.acm.org/citation.cfm?id=1241550.

MacKay, B., Kellar, M., & Watters, C. (2005). An evaluation of landmarks for re-finding information on the web. *CHI '05: CHI '05 extended abstracts on Human factors in computing systems* (pp. 1609-1612). New York, NY, USA: ACM. doi: http://doi.acm.org/10.1145/1056808.1056978.

MacKay, B., & Watters, C. (2008). Exploring multi-session web tasks. *CHI '08: Proceeding of the twenty-sixth annual SIGCHI conference on Human factors in computing systems* (pp. 1187-1196). New York, NY, USA: ACM. doi: http://doi.acm.org/10.1145/1357054.1357243.

Malone, T. W. (1983). How do people organize their desks?: Implications for the design of office information systems. *ACM Transactions on Information Systems (TOIS)*, *1*(1), 99-112. Retrieved from http://portal.acm.org/citation.cfm?id=357430.

Morrison, J. B., Pirolli, P., & Card, S. K. (2001). A taxonomic analysis of what world wide web activities significantly impact people's decisions and actions. *Conference on Human Factors in Computing Systems* (pp. 163-164). New York, NY, USA: SIGCHI: ACM Special Interest Group on Computer-Human Interaction. Retrieved from http://portal.acm.org/citation.cfm?id=634167.

Obendorf, H., Weinreich, H., Herder, E., & Mayer, M. (2007). Web page revisitation revisited: implications of a long-term click-stream study of browser usage. *CHI '07: Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 597-606). New York, NY, USA: ACM. doi: http://doi.acm.org/10.1145/1240624.1240719.

Robertson, G., Czerwinski, M., Larson, K., Robbins, D. C., Thiel, D., & Dantzich, M. van. (1998). Data mountain: using spatial memory for document management. *UIST '98: Proceedings of the 11th annual ACM symposium on User interface software and technology* (pp. 153-162). New York, NY, USA: ACM. doi: http://doi.acm.org/10.1145/288392.288596.

Sawasdichai, N., & Poggenpohl, S. (2002). User purposes and information-seeking behaviors in web-based media: a user-centered approach to information design on websites. *DIS '02: Proceedings of the 4th conference on Designing interactive systems* (pp. 201-212). New York, NY, USA: ACM. doi: http://doi.acm.org/10.1145/778712.778742.

Sellen, A. J., Murphy, R., & Shaw, K. L. (2002). How knowledge workers use the web. *CHI '02: Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 227-234). New York, NY, USA: ACM. doi: http://doi.acm.org/10.1145/503376.503418.

Sugiura, A., & Koseki, Y. (1998). Internet scrapbook: automating Web browsing tasks by demonstration. *UIST '98: Proceedings of the 11th annual ACM symposium on User interface software and technology* (pp. 9-18). New York, NY, USA: ACM. doi: http://doi.acm.org/10.1145/288392.288395.

Tauscher, L., & Greenberg, S. (1997a). Revisitation Patterns in World Wide Web Navigation. *CHI '97: Proceedings of the SIGCHI conference on Human factors in computing systems* (pp. 399-406). New York, NY, USA: ACM. doi: http://doi.acm.org/10.1145/258549.258816.

Tauscher, L., & Greenberg, S. (1997b). How people revisit web pages: empirical findings and implications for the design of history systems. *International Journal of Human-Computer Studies*, *47*(1), 97-137. doi: 10.1006/ijhc.1997.0125.

Teevan, J. (2005). The Re:Search Engine - Helping People Return to Information on the Web. Seattle, WA: In Proceedings of the ACM Symposium on User Interface Software and Technology (UIST '05).

Teevan, J., Alvarado, C., Ackerman, M. S., & Karger, D. R. (2004). The perfect search engine is not enough: a study of orienteering behavior in directed search. *Conference on Human Factors in Computing Systems* (pp. 415 - 422). Vienna, Austria: ACM. Retrieved from http://portal.acm.org/citation.cfm?id=985745.

Weinreich, H., Obendorf, H., Herder, E., & Mayer, M. (2008). Not quite the average: An empirical study of Web use. *ACM Trans. Web*, *2*(1), 1-31. New York, NY, USA: ACM. doi: http://doi.acm.org/10.1145/1326561.1326566.

White, R. W., & Drucker, S. M. (2007). Investigating behavioral variability in web search. *WWW '07: Proceedings of the 16th international conference on World Wide Web* (pp. 21-30). New York, NY, USA: ACM. doi: http://doi.acm.org/10.1145/1242572.1242576.

Whittaker, S., & Hirschberg, J. (2001). The character, value, and management of personal paper archives. *ACM Transactions on Computer-Human Interaction (TOCHI)*, *8*(2), 150-170. Retrieved from http://portal.acm.org/citation.cfm?id=376932.

Whittaker, S., & Sidner, C. (1996). Email overload: exploring personal information management of email. *Conference on Human Factors in Computing Systems* (pp. 276 - 283). New York, NY, USA: ACM. Retrieved from http://portal.acm.org/citation.cfm?id=238530.

APPENDIX A   FORMS

A.1 INFORMED CONSENT FORM

Informed Consent Form – *Please read carefully before signing*

DALHOUSIE UNIVERSITY
Inspiring Minds

Interface Models to Manage Information Gathering on the Web
Principal Investigators: Timothy Mankowski, Computer Science Graduate Student
                            Dr. Carolyn Watters, Faculty of Computer Science
Contact Person: **Timothy Mankowski, Faculty of Computer Science, timothym@cs.dal.ca**

We invite you to take part in a research study being conducted by Timothy Mankowski who is a graduate student at Dalhousie University. Your participation in this study is voluntary and you may withdraw from the study at any time. Your academic (or employment) performance evaluation will not be affected by whether or not you participate. The study is described below. This description tells you about the risks, inconvenience, or discomfort which you might experience. Participating in the study might not benefit you, but we might learn things that will benefit others. You should discuss any questions that you have about this study with Mr. Mankowski.

The purpose of this study is to learn how to design better Web browser tools for personal computers (i.e.: desktops and laptops). You will be asked to participate in one session at the Faculty of Computer Science for which you will be compensated $15.

During the session you will be first asked to fill in a background questionnaire detailing your previous experience performing tasks using a web browser. You will then be given the first of two short tutorials explaining how to use a customized version of the Mozilla Firefox Web browser. Following the first tutorial is a 15 minute task on the Web during which you will use the customized Web browser to gather information from websites. After the first task is complete, you will be given a post-task questionnaire. A second tutorial will follow, along with another task and questionnaire. A final third questionnaire is given once each of the two tutorials, tasks and questionnaires are completed. The last stage of the user study is the post-study interview. The interview will assess both tasks performed in the first session, your overall opinions and comments.

A researcher will be available at all times to answer any questions might have with regard to the tasks or software being used during the study. The customized browser will log your user interactions. We ask that you please refrain from checking any password protected accounts (e.g.: Webmail, Facebook, or Twitter) during the sessions. All personal and identifying data will be kept confidential. Anonymity of textual data will be preserved by using pseudonyms. The informed consent form and all research data will be kept in a secure location under confidentiality in accordance to Dalhousie University policy for 5 years post publication.

In the event that you have any difficulties with or wish to voice concern about any aspect of your participation in this study, you may contact Patricia Lindley, Director, Office of Research Ethics Administration at Dalhousie University's Office of Human Research Ethics for assistance: phone: (902) 494-1462, email: patricia.lindley@dal.ca.

"I have read the explanation about this study. I have been given the opportunity to discuss it and my questions have been answered to my satisfaction. I hereby consent to take part in the study. However, I understand that my participation is voluntary and that I am free to withdraw from the study at any time."

Participant   *Researcher*

Name: _____   Name: _____
Signature: _____   Signature: _____
Date: _____   Date: _____

Page 1 of 2

Please select <u>one</u> of the options below:

"I consent to the direct quotation of any comments or statements made in any written reports without viewing the quotes prior to their use and I understand that the anonymity of textual data will be preserved by using pseudonyms."

Participant   **Researcher**

Name: _____   Name: _____

Signature: _____   Signature: _____
Date: _____   *Date:* _____

<u>*or*</u>

"I want to read direct quotes prior to their use in reports and I understand that the anonymity of textual data will be preserved by using pseudonyms."
[if this option is chosen, please include a contact email address and phone number:
_____ _____]

Participant   **Researcher**

Name: _____   Name: _____

Signature: _____   Signature: _____
Date: _____   *Date:* _____

A.2 BACKGROUND QUESTIONNAIRE

Background Questionnaire
Please complete the information below

<div style="text-align:right">

User ID

</div>

PART I - PLEASE FILL IN THE FOLLOWING INFORMATION:

1. Age: _____

2. Gender: ☐ Male ☐ Female

3. Level/Year:
☐ 1st Year Undergraduate ☐ 2nd Year Undergraduate
☐ 3rd Year Undergraduate ☐ 4th Year Undergraduate
☐ Graduate – Masters ☐ Graduate – PhD
Other _____

4. a) How often do you use the Web:

☐ Several Times a Day (5 or more)
☐ Few Times a Day (3-5 times)
☐ Daily (at least once)
☐ Several Times a Week (5 or more)

4. b) How many hours (approximately) do you spend a week using the Web:

☐ 1-2 hours
☐ 3-5 hours
☐ 6-10 hours
☐ 11-15 hours
☐ 16-20 hours
☐ 20+ hours

5. Where do you primarily use the Web:
☐ At School
☐ At Home
☐ At Work
☐ No one place - use the Web equally across locations (e.g. you have a laptop)
☐ Other _____

PART II - PLEASE FILL IN THE FOLLOWING INFORMATION:

6. What web browsers have you ever used? (Please check all that apply)
☐ Internet Explorer ☐ Opera ☐ Google Chrome
☐ Netscape ☐ Firefox ☐ Other _____
☐ Mozilla ☐ Safari

7. Currently what is your primary web browser? (Please check one)
☐ Internet Explorer ☐ Safari
☐ Firefox ☐ Google Chrome
☐ Opera ☐ Other _____

8. How long has this been your primary browser?

Page 1 of 2

9. Below are different browser tools and actions. *Please circle* your usage of each tool

Tools/Actions	1 Never heard of	2 Never Use	3 Use infrequently	4 Use Sometimes	5 Use often
Open more than one browser window	1	2	3	4	5
Use tabs	1	2	3	4	5
Use search engines	1	2	3	4	5
Search within a web page	1	2	3	4	5
Use Bookmarks/ Favorites	1	2	3	4	5
Email Information to myself or others	1	2	3	4	5
Use History Lists	1	2	3	4	5
Save web pages	1	2	3	4	5
Print web pages	1	2	3	4	5
Type a URL from memory	1	2	3	4	5
Use Auto complete	1	2	3	4	5
Make notes about information I see (paper or electronic)	1	2	3	4	5
Copy/Paste information to or from another application (e.g., Word, Notepad…)	1	2	3	4	5
Save sessions	1	2	3	4	5
Other Tool (please add):	1	2	3	4	5
Other Tool:	1	2	3	4	5
Other Tool:	1	2	3	4	5
Other Tool:	1	2	3	4	5

10. Do you use the Web to gather information or special topics?

11. How would you **start** if you were planning a trip to Paris?

A.3 POST-TASK QUESTIONNAIRE

Post-task Questionnaire

Please *circle* the appropriate answer to each question, after reading <u>each</u> question *carefully*.

User ID	Task ID	Tool ID

	1 Strongly disagree	2 Disagree	3 Neither agree nor disagree	4 Agree	5 Strongly agree
This tool was easy to use.	1	2	3	4	5
This tool was **hard** to learn.	1	2	3	4	5
I would recommend this tool to other people.	1	2	3	4	5
I would like to see this tool as a standard feature in all Web browsers.	1	2	3	4	5
This tool contains useful new features.	1	2	3	4	5
Why or why not?					
This tool helped me complete my task.	1	2	3	4	5
I would use this tool for more challenging tasks.	1	2	3	4	5
Why or why not?					
I would use this tool for tasks that take longer to complete.	1	2	3	4	5
For example:					
This tool let me compare information from different websites.	1	2	3	4	5
I think this tool would help me remember information or websites.	1	2	3	4	5
How?					
This tool was useful for remembering important **text**.	1	2	3	4	5
Why or why not?					
This tool was useful for remembering important **images**.	1	2	3	4	5
Why or why not?					
I needed more time to complete this task.	1	2	3	4	5
I have performed this type of task on the Web before.	1	2	3	4	5

Page 1 of 2

Any comments?

A.4 POST-SESSION SURVEY

Post-task Survey

Please *circle* the appropriate answer to each question.
Please read each question *carefully*.

User ID

	1 Strongly disagree	2 Disagree	3 Neither agree nor disagree	4 Agree	5 Strongly agree
Page Thumbnails was easier to use than Web Scraps.	1	2	3	4	5
I enjoyed using Web Scraps more than Page Thumbnails.	1	2	3	4	5
Page Thumbnails was more helpful for complex tasks than Web Scraps.	1	2	3	4	5
Why or why not?					
Page Thumbnails was easier to learn than Web Scraps.	1	2	3	4	5
Why or why not?					
Web Scraps helped me complete my task more efficiently than Page Thumbnails.	1	2	3	4	5
I think Page Thumbnails helped me remember more relevant information than Web Scraps.	1	2	3	4	5
Web Scraps helped me compare information between websites more easily than Page Thumbnails.	1	2	3	4	5
I think I can recall *specific information* easier by looking at Page Thumbnails than Web Scraps.	1	2	3	4	5
Why or why not?					
I think I can recall *entire websites* easier using Page Thumbnails than Web Scraps.	1	2	3	4	5
Why or why not?					
I think Web Scraps helped me remember why I chose to keep something more efficiently than Page Thumbnails.	1	2	3	4	5
Why or why not?					

A.5 POST-TASK SEMI-STRUCTURED INTERVIEW

Post-Task Semi-Structured Interview

User ID

Your task question was:

☐ Task A – *Day Trip*

☐ Task B – *Golf Trip*

The tool that you used was:

☐ Tool X – *Page Thumbnails*

☐ Tool Y – *Layered Web Scraps*

1. a. What strategies did you take while performing this task?

   b. Did you leave a lot of tabs or windows open? Why?

2. Why did you choose to keep these websites/selections?

3. Are there any new features in this tool that you found useful? Why?

4. Can you think of any other features that might have helped you to complete your task more easily or make this tool more useful?

5. Any comments before we move on to the next section?

_____

A.6 POST-STUDY SEMI-STRUCTURED INTERVIEW

Post-Study Semi-Structured Interview

User ID

1. Was one tool more useful than the other? Why? In which situations?

2. Which tool did you like better? Why?

3. Did the tasks allow you to make proper use of both tools?

4. Were there enough boxes to gather everything that you wanted to gather?

5. What are your overall thoughts?

APPENDIX B   TASKS

B.1 TASK A – DAY TRIP

You are working for a daycare centre in Bedford, N.S. You have been asked to help organize a day trip (9:30 am to 12:30 pm) around downtown Halifax for a group of 18 kids ages 6 to 10 and two adults. **Use the tool demonstrated to collect information for your daycare supervisor.**

Your task is to:

    A. recommend 2 suitable places to see during the hours of the outing;

    B. find prices for snacks and juice for everyone, and;

    C. find prices for public transit bus tickets and rates for a local taxi in case of an emergency.

Your daycare supervisor wants to get an idea of how much food and transport will cost, as a separate item each (**<u>do not</u>** calculate a total price or how many items to purchase), see below for further instructions.

<u>Summary</u>:

- Find Metro Transit **bus** ticket prices for adults and children, and the cheapest fare for a taxi from the places of interest back to the daycare in Sunnyside Mall in Bedford.

- Find appropriate **snacks** (no peanuts) and **juice** to fill the drink cooler (see picture); take note of prices. Students will bring their own lunches.

- **2 interesting places** to see between 9:30 am and 12:30 pm (do not worry about costs here)

B.2 TASK B – GOLF TRIP

You're working at ABC Corp in Halifax, N.S. You've been asked to help organize the company golf trip to Kingston, Ontario. **Use the tool demonstrated to collect information for your boss.**

Your task is to:

A. Find prices for a golf umbrella and a golf ball (any colour other than white);

B. Recommend two fine dining restaurants in Kingston for all 18 employees (check whether they offer champagne), and;

C. Get a price estimate for a taxi and an airport shuttle from the Lord Nelson Hotel to Halifax International Airport. Your boss wants to know how much the umbrellas, golf balls, and airport transportation will cost (**do not** calculate a total price or how many items/services to purchase, see below for further instructions).

Summary:

- Find prices to the airport for **1 taxi company** and **1 bus shuttle** company that depart from the Lord Nelson Hotel

- Find prices for 1 **golf umbrella** and 1 **golf ball** (any colour other than white)

- Suggest **2 fine dining restaurants** in Kingston, Ontario (check whether they offer champagne)

APPENDIX C   QUALITATIVE DATA

C.1 POST-TASK QUESTIONNAIRES

Table 14: Qualitative Data from Post-Task Questionnaires

Question	Tool	Comment (Count)
5. This tool contains useful new features.	WS	Selecting specific information/areas (12) Combining information into one place (4) Enjoyed the tool's efficiency (4) Comparing information (2) Not really useful (1)
	PT	Used or seen this tool before (9) Not really useful (7) More visual than using tabs or bookmarks (6)
7. I would use this tool for more challenging tasks.	WS	Useful for more complex tasks (14) Not useful for more complex tasks (6) Uncertain whether useful for more complex tasks (4) Useful for saving specific information (4) Helped remember why information was saved (4) Useful for comparison (3)
	PT	Not useful for more complex tasks (9) Useful for more complex tasks (7) Uncertain whether useful for more complex tasks (6) Useful for saving/remembering information (5) Six Page Thumbnails not enough (2)
8. I would use this tool for tasks that take longer to complete	WS	Useful for longer tasks (14) Uncertain whether useful for longer tasks (6) Not useful for longer tasks (3) Useful for comparing information (4) Six Webscraps not enough (3) Remember why information was saved (3) Saving specific information (6) More useful for shorter tasks (3)

Question	Tool	Comment (Count)
	PT	Useful for longer tasks (10) Would use for … (9) Uncertain whether useful for longer tasks (4) Not useful for longer tasks (4)
10. I think this tool would help me remember information or websites.	WS	Useful (18) Capturing specific information (11) Comparing information (4) Focusing on web pages (3) Possibly useful (3) Not useful (2)
	PT	Remembering websites (15) Useful (15) Does not save specific information (6) Not useful (6) Similar to bookmarking (4) Possibly useful (4)
11. This tool was useful for remembering important text.	WS	Selecting/highlighting specific area/text (15) Prefer using other methods to remember text (3)
	PT	Useful for remembering entire web pages (8) Not useful for remembering text (7) Text within web pages was too small to read (7)
12. This tool was useful for remembering important images.	WS	Useful for selecting images (11) Gathered few or no images (4)
	PT	Pictures are easily viewable within Page Thumbnails (10) Gathered few or no images (7)

C.2 POST-SESSION SURVEY

Table 15: Qualitative Data from Post-Session Surveys

Question	Tool	Comment (Count)	+	N	–
3. Page Thumbnails was more helpful for *complex tasks* than Web Scraps	WS	Helps to focus on specific information (14) No page re-visitation required (3) Helps to compare information (2)	20	4	0
	PT	Easier to use (2)	3		1
4. Page Thumbnails was *easier to learn* than Web Scraps	WS	Has more features (7)	0	12	2
	PT	Requires fewer steps (5)	6		0
	Both	Easy to use (3) Similar difficulty (8)	-		-
8. I think I can *recall specific information* more easily by looking at Page Thumbnails than Webscraps	WS	Highlight important parts of a web page (12) Easier to remember information (3) Easier to read (2)	16	2	0
	PT	-	2		3
9. I think I can *recall entire websites* easier using Page Thumbnails than Webscraps	WS	-	3	2	1
	PT	Seeing the entire page helps recall websites (10) Seeing irrelevant information helps with web page recall (2)	14		0
10. I think that Webscraps helped me remember *why I chose to keep something* more efficiently than Page Thumbnails	WS	Highlights/focuses information (16)	19	3	0
	PT	-	0		0
Webscraps Total			**58**	**23**	**3**
Page Thumbnails Total			**14**		**4**

C.3 POST-TASK SEMI-STRUCTURED INTERVIEWS

Table 16: Qualitative Data from Post-Task Semi-Structured Interviews

Question		
Comment	**Count**	
	WS	**PT**
1A. What strategies did you make while performing this task?		
Used a traditional text-based search engine such as Google	24	23
Opened search results within several new tabs or windows	12	10
Added new terms/parameters to improve search results	5	3
Used a different search engine (e.g., Google maps/images, eBay)	3	5
Searched within or navigating within pages or websites	3	3
Reused the same window for searches / using the back button in searches	2	2
1B. Did you leave a lot of tabs or windows open?		
Information was still useful or needed more time to read information	16	9
It was confusing to keep many tabs open	5	3
Too lazy / Forgot to close tabs	4	4
Kept tabs open for comparison	3	1
2. Why did you choose to keep these websites/selections?		
To answer the question	17	20
They were the first things I found	2	4
Cheapest or contained the price of an item/service	5	4
I encountered software bugs while using this tool	2	0
3. Are there any new features in this tool that you found useful?		
Selecting specific information	17	0
Getting back to information is easier	7	0
Comparing and clustering	6	5

Question		
Comment	**Count**	
	WS	**PT**
Criticism / room for improvement	6	3
Useful for particular tasks	5	2
Helping to remember information / reminding	3	3
Background in Webscraps was useful	3	N/A
Useful to represent pages visually	0	10
Similar to another tool/I have used this tool before	0	6
4. Can you think of any other features that might have helped you complete your task more easily or make this tool more useful?		
More than 6 Webscraps/Page Thumbnails per page / Custom number	13	9
Be able to save categories of Webscraps/Page Thumbnails	9	3
Be able to re-size, zoom, or scroll	6	13
Did not use Webscraps' background or did not like it	5	N/A
Suggestions/Comments for other/additional user interface controls	4	4
Clicking directly on a Webscrap/Page Thumbnail to open a link (Instead of clicking on the text hyperlink for WS/PT)	3	8
Re-ordering/Arranging Webscraps/Page Thumbnails	3	3
Searching within Webscraps/Page Thumbnails	2	2
Implement custom attributes (e.g., text annotations, custom titles)	1	4

C.4 POST-STUDY SEMI-STRUCTURED INTERVIEW

Table 17: Qualitative Data from Post-Study Semi-Structured Interviews

Question	
Webscraps (Count)	**Page Thumbnails (Count)**
1. *Was one tool more useful than the other? Why? In which situations?*	
WS better (14): comparison situations (2), focusing on information (10) **It depends/both good (9)** *Why*: focus on information (5), Good for complex tasks (4), Useful for comparison (3) **Both tools similar/same (4)**	**PT better (2)**: more user friendly/easier to use (2) *Why*: Web page screen shots/saving whole pages (2), For simpler tasks (2)
2. *Which tool did you like better? Why?*	
WS better (18): focus/concentrate on specific info (4) **Neither tool (2)**	**PT better (7)**: simpler interface (3)
3. *Did the tasks allow you to make proper use of both tools?*	
Yes (20): gathered information (3), compared information (2) **Yes, but… (5)** **No (5)**: Didn't get to perform comparisons (2)	
4. *Were there enough boxes to gather everything that you wanted to gather?*	
Yes, but… (13): Enough for both tasks, but not enough for practical purposes (7) **Yes (6)** **No (5)**: Not enough for both tasks (5) **Yes and no (4)**: one of the tasks okay, needed more boxes for the other task (4)	

C.5 SUMMARY OF QUALITATIVE DATA

Table 18: Summary of Qualitative Data

Rank	Webscraps Comment	Count	Page Thumbnails Comment	Count
1	Selecting specific information/areas	127	Remembering websites	39
2	Comparing information	26	Similar to another tool / I have used this tool before	16
3	Why information was saved	16	Easier to use than Webscraps	15
4	Useful for example…	14	Wanted: re-sizing, zooming, and scrolling	13
5	No page re-visitation required	13	Useful for example...	12
6	Wanted: categories of Webscraps	12	Pictures are easily viewable within Page Thumbnails	10
7	Useful for selecting images	11	Represent pages visually	10
8	Combining information into one place	10	Wanted: clicking directly on image instead of URL	8
9	Layer (Background) was helpful	7	Comparing information	7
10	Layer (Background) was not helpful	6	Gathered few or no images	7
11	Negative aspects / Room for improvement	6	Text too small to read	7
12	Wanted: re-sizing, zooming, and scrolling	6	Did not help to remember text	7
13	Comparing prices	5	New features were not useful	7
14	Easier to read than Page Thumbnails	5	Does not save specific information	6
15	Enjoyed the tool's efficiency	5	More visual than using windows, tabs or bookmarks	6

Rank	Webscraps Comment	Count	Page Thumbnails Comment	Count
16	Remembering websites	5	Saving/Remembering information	6
17	Gathered few or no images	4	Similar to bookmarks	5
18	New idea	4	Wanted: categories of Page Thumbnails	5
19	Wanted: controls (interface buttons, hotkeys, etc.)	4	Wanted: controls (interface buttons, hotkeys, etc.)	4
20	Did not do much comparison	3	Wanted: custom Attributes (e.g., text annotations, custom titles, relationships)	4
21	More features	3	Helping to remember information / reminding	3
22	Preferred other methods to remember text	3	Negative aspects / Room for improvement	3
23	Useful for shorter tasks	3	Wanted: re-ordering/arranging boxes	3
24	Wanted: clicking directly on image instead of URL	3	Wanted: search within selections	2
25	Wanted: re-ordering/arranging boxes	3	Seeing irrelevant info helps with page recall	2
26	Complex information or tasks	2	Useful for simple tasks	2
27	Wanted: search within selections	2	Less clutter than Webscraps	1
28	New features were not useful	1	No page re-visitation required	1
29	Useful for simple tasks	1	Helped remember why information was saved	1
30	Wanted: custom Attributes (e.g., text annotations, custom titles, relationships)	1	**TOTAL**	**212**
	TOTAL	**311**		