Studying the User Task of Information Gathering on the Web

by

Anwar Alhenshiri

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy

at

Dalhousie University
Halifax, Nova Scotia
March 2013

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 "Studying the User Task of Information Gathering on the Web" by Anwar Alhenshiri in partial fulfilment of the requirements for the degree of Doctor of Philosophy.

Dated: March 13, 2013

External Examiner: ______________________________

Research Supervisor: ______________________________

Examining Committee ______________________________

____

Departmental Representative: ______________________________

DALHOUSIE UNIVERSITY

DATE: March 13, 2013

AUTHOR: Anwar Alhenshiri

TITLE: Studying the User Task of Information Gathering on the Web

DEPARTMENT OR SCHOOL: Faculty of Computer Science

DEGREE: PhD CONVOCATION: May YEAR: 2013

Signature of Author

DEDICATION

This dissertation is dedicated to the first two people I loved, my parents. May my father rest in peace!

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

ABSTRACT

Research has studied information seeking behaviour and several models have been developed. Those models were further modified following the emergence of the web. At the beginning of the 2000s, research started focusing on the concept of a user task instead of an activity or a simple action. The studies conducted were aimed to categorize the user activities into high level tasks. Investigating the tasks identified is anticipated to assist with developing tools and applications that would help the user to accomplish those tasks.

After categorizing the user information seeking activities into high-level tasks, research continued to investigate changes in the frequencies of the tasks identified. Changes in the user behaviour that accompanied the evolution of the web and its applications have been targeted for improving how users interact with tools intended for accomplishing user activities. However, there has been little emphasis on studying the high-level tasks identified in the case of the web. Even though those tasks differ substantially, users have been using the same web browsing model to accomplish most of the activities under each type of task.

The research discussed in this dissertation is concerned with studying the task of information gathering which is also known as the informational task. This task was selected due to: 1) its high frequency on the web (between 48% and 61.25% of the overall tasks users perform); 2) its complexity and the ambiguity associated with the kind of activities that comprise the task; 3) the need for using multiple applications for accomplishing the requirements of this type of task; and 4) the necessity for collecting different types of data from different sources for satisfying the task requirements.

The current state in research related to information gathering identifies this task based on a simple description of the user activities that distinguishes information gathering from other kinds of tasks. The research discussed in this dissertation: 1) provides a thorough definition of the task, 2) models its underlying subtasks (sets of related activities), and 3) investigates difficulties and issues associated with each subtask. The investigations lead to design recommendations that resulted in building specific features to be examined during information gathering tasks. The research concluded by providing final recommendations based on the findings which resulted from investigating those features.

LIST OF ABBREVIATIONS USED

HTML	Hyper Text Markup Language
XML	eXtensible Markup Language
VSE	Visual Search Engine
VLN	Visual Link Navigation Interface
WIGI	Web Information Gathering Interface
PDF	Portable Document Format

ACKNOWLEDGEMENTS

I thank my committee members who guided me throughout my journey to complete my PhD. Special thanks to my wife who always supported me. Thanks to all the participants in all of the studies involved in this research. Thanks to the Faculty of Computer Science at Dalhousie University. I also thank my twins who delayed me as much as they could.

CHAPTER 1 INTRODUCTION

Prior to the emergence of the web, several information seeking models were introduced in the literature. Those models showed the sequences of activities users perform to locate the information needed. However, those models were concerned with identifying the user activities and their underlying actions without considering the broader understanding of the user task or goal. Following the emergence of the web, research started to focus more on the concept of a complete task by identifying the kinds of tasks users perform on the web as a first step.

A task implies more than one activity and involves targeting a broader goal. Different models have been built to categorize the types of tasks identified. In the work of Broder (2002), the research identified different kinds of tasks including: navigational, informational, and transactional tasks. Similarly, but with different labeling of the tasks identified, Sellen (2002) categorized user tasks on the web into: information gathering, finding, browsing, transacting, communicating, and housekeeping. Following on the work of Broder (2002), Rose and Levinson (2004) identified user goals based on the tasks being accomplished as: navigational, informational, and resources. Kellar, et al. (2007) developed a model of user tasks on the web and identified: fact finding, information gathering, transacting, and browsing as the main kinds of tasks users perform based on the results of a field study. In these models, informational tasks were shown to be very common on the web.

After categorizing the kinds of tasks users perform on the web, research attempted to examine some of those tasks further. The task of information gathering which represents between 48% (Broder, 2002) and 61.25% (Rose and Levinson, 2004) of all the tasks users perform on the web was further investigated in the work of Amin (2009). The research in Amin (2009) was an attempt to identify some of the characteristics of the task of information gathering. Information gathering (also labeled as informational) tasks were shown to be complex, highly search reliant, and to involve different types of search including: exploratory, comparison, and topic search.

Even though the kinds of tasks users perform on the web have been identified and categorized, research has made little progress in terms of studying each individual task.

Categorizing user tasks on the web was in one way intended to shift the research and investigations from looking into individual activities to studying broader tasks. The task of information gathering on the web was selected for investigation in this research because of the following problems and issues associated with this type of task:

1- Information gathering implies collecting information from different sources and comparing information for decision making.
2- Information gathering requires the use of multiple applications and tools. The effectiveness of current web tools with regard to how users accomplish the task of information gathering has not yet been investigated.
3- Information gathering is a complex task that consists of underlying subtasks which research has not yet identified.
4- Information gathering is very common on the web.

To study the task of information gathering, different investigations were carried out in this research. The research discussed throughout this dissertation attempts to answer the following questions:

1- What are the subtasks that comprise the task of information gathering in the case of the web? How can they be modeled?

2- What is the definition of the task of information gathering on the web?

3- What are the specific features that distinguish the task of information gathering from other types of tasks? How do they contribute to the subtasks identified?

4- To be used in investigations, how can tasks of information gathering be built to simulate realistic tasks?

5- How do users currently perform the subtasks of the information gathering task? What tools do they use and what difficulties do they encounter?

6- What features should be built and investigated to examine their effectiveness in tools intended for information gathering?

7- What practical design recommendations can be developed?

To study and investigate the task of information gathering, this research has gone through several steps. First, reviewing the literature helped with creating a model in which the subtasks comprising the task of information gathering on the web were initially identified. The model is shown in Figure 9. Those subtasks are combinations of activities that have the same goal. The model helped with conducting two studies in which the subtasks of finding information sources (web pages and sites) and finding information were investigated in the second and third steps of this research.

A small-scale study (Alhenshiri, et al., 2010b) was conducted to explore the use of visualization and visual clustering for gathering information on the web as discussed in Chapter 4. The study was meant to investigate different visualization aspects in improving web search for information gathering. Searching for gathering information by the submission and resubmission of search queries to search engines is usually adopted by users gathering information on the web (Teevan, et al., 2004). The results of the study showed that users benefited from visualizing features of web documents in finding sources of information that covered more topics related to the task at hand. The study highlighted the need for exploring how users follow links on the web graph to find more information for the task and compare information for decision making.

The third step in the research was conducting a study (Alhenshiri, et al., 2010a) to follow up on the recommendations from the previous step. The study concerned gathering information through searching by following link hierarchies on the web graph, an approach used by users of the orienteering search behaviour discussed in the work of Teevan, et al. (2004). The subtasks investigated in this study were finding information and finding further related information as shown on the initial model in Figure 9. The study further highlighted the complexity of the task of information gathering on the web and showed the effectiveness of several visualization features on the search process for information gathering. The features implemented to assist with gathering information helped with navigating through websites and achieved more user confidence in the task results.

Following the two preliminary studies, the information gathering task model was modified as shown in Figure 22. The fourth step in this research explored the subtask of information management and organization and the subtask of re-finding information on the web. The subtasks chosen for investigation were decided after conducting a questionnaire-based study discussed in Alhenshiri, et al. (2011). In that study, users indicated difficulties and concerns regarding those two subtasks in particular. The exploratory study (Alhenshiri, et al., 2012a) used simulations of information gathering tasks to explore how users accomplish those tasks on current tools and applications available and to develop recommendations for further studies. The study showed that the subtasks of re-finding information and managing and organizing information required further support in current web tools. In addition, the subtask of handling multiple sessions was added to the model of the task of information gathering to be considered for further investigations as shown in Figure 32. The study is discussed in Chapter 5.

The fifth and last step in this research was a study (Alhenshiri, et al., 2012b and 2012c) conducted based on the recommendations developed in the previous step. The recommendations involved the subtasks of managing and organizing information, handling multiple sessions, and re-finding information while performing information gathering tasks on the web. The study involved building a prototype system in which specific features were embedded and their effectiveness in how users performed the tasks was measured.

The results of the final study showed that keeping track of user references to sources of information in a specific area on the display using thumbnails of web pages achieved a significant difference over the different strategies users use to track sources of information. Moreover, allowing the user to keep the task information integrated and later restart the whole task as one unit without losing the task context was significantly more effective than the use of other strategies available with current tools. Finally, the study showed that adding the abilities to edit and format the information the user gathers to browsing and searching for information was more effective than the use of multiple tools and applications. The study is discussed in Chapter 6.

The research used some guidelines from the work of (Kules and Capra, 2008) to develop simulations of realistic user tasks for information gathering on the web. Other principles for creating the tasks were also developed in this research as discussed in Chapter 5. The principles developed here were used to ensure that the tasks created were at similar levels of complexity. Those principles can be used in investigating the task of information gathering with future tools and features built for improving the effectiveness of this type of task. The principles can also be used for developing and investigating other kinds of tasks.

This research investigated the concept of information gathering to provide a more robust definition of this type of task since the current definitions are based only on simple descriptions. The information gathering task was defined as a combination of multiple subtasks each of which consists of different activities. Information gathering is complex and highly search-reliant. It requires collecting different kinds of data possibly from different sources. It may necessitate the use of multiple applications, and it may require multiple sessions to complete.

The research resulted in developing a model of the task of information gathering that went through three different stages. An initial model was created based on the literature review. As a result of the investigations that examined certain activities related to information gathering, the initial set of subtasks was developed and modeled. The initial model was modified in accordance with the results of the first two studies as shown in Figure 22. The changes involved the subtasks in the model. In addition, the features associated with each subtask were added to the model after defining the task of information gathering.

After conducting an exploratory study that examined the current state of information gathering in terms of how users currently use available tools and strategies to perform information gathering tasks, the model was further modified as shown in Figure 32. The subtask of handling multiple sessions was added to the model as recommended in the results of the study. The final model clarified the components of the task of information gathering on the web as demonstrated in this research. The final set of core subtasks were shown to be: finding information sources, finding information, managing and organizing information, re-finding information, and handling multiple sessions.

The model developed in this research can be used in further investigations to the subtasks that were not covered in this research such as comparing information, interpreting the task, and reviewing the task. Subtasks may be added, removed, or modified in future research. The model depicts the current process of how users gather information from the web. The tools and features needed for each subtask can be further enhanced by continuing the investigations on this model. The model can also help with studying other types of tasks on the web such as fact finding and browsing by analyzing those subtasks using the same approach followed in this research.

Future research can also use the model discussed in this dissertation and the features investigated to study information gathering from the web on small-screen devices such as tablets and smart phones. The increasing demand on the use of those devices especially for web tasks may necessitate investigating the kinds of tasks users perform on those devices. Information gathering in particular is a task that may require building and investigating different features in the case of small-screen devices from those investigated in this research. Future research may also benefit from the current model and the task creation process in studying information gathering on small-screen devices.

CHAPTER 2 RELATED WORK

The research discussed in this chapter reviews: information seeking models, models of user tasks on the web, and investigations related to aspects of the information gathering task on the Web. In addition, the discussion covers aspects of visualization and clustering examined in relationship to how users gather information from the web.

2.1. INFORMATION SEEKING MODELS

In order to categorize user activities on the web, researchers often look to models of information seeking (Ellis, 1989; Marchionini, 1995; Choo, et al., 1998, 2000). Many of the models discussed in the literature, although help with understanding the user behaviour in certain activities, lack several features related to particular aspects of the web. The following discussion covers some of the most well-known information seeking and user behaviour models. The activities which apply to the case of the web are further discussed where applicable.

2.1.1.Ellis's Model

Ellis (1989) stated that there are six main activities applicable to hypertext environments (of which the web is one). Those activities represent user actions during seeking information that is not previously known to the user and which is aimed to increase the user's knowledge. The boundaries among those activities are rather soft. The following is a summary of user activities in Ellis' Model, which are shown in Figure 1.

Starting. Starting is selecting a starting point for searching. For example, on the web, starting could be as simple as going to a website such as Yahoo!'s to find a category or submitting a search query to a search engine. Selecting a starting point could also be by selecting a bookmarked page. In Ellis' (1997) modified model, starting is called *Surveying*. Surveying is the activity of finding key people or key resources in the field of information being sought.

Chaining. This activity means following references from a starting point such as following references on an article in the aim of finding further information. Chaining can happen either backward or forward. For example, using a web browser, forward chaining would be following links on the starting page whereas backward chaining could be by using a search engine to find web pages that link to the starting page.

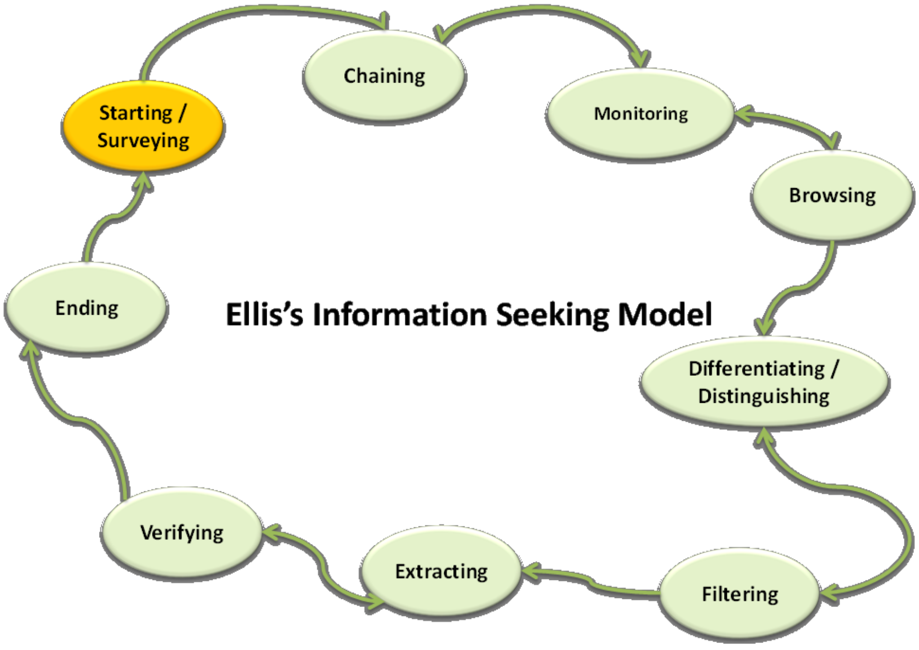


Figure 1. Ellis' model.

Monitoring. That is observing a source of information for changes. An example can be going back to a web page frequently to see if any updates have occurred on the page or a new edition of the page has been published.

Browsing. Browsing is navigating through documents by following links in different directions in the case of the web. This process can result in establishing several starting points and it is open to serendipitous findings. Following the table of contents or the site structure on the web are examples of browsing. An alternate term that is commonly used for browsing is navigation.

Differentiating. Choosing sources of information for the task is called differentiating according to this model. The process of differentiating depends on the experience of the information seeker and the relatedness of the source to the task's topic. A web example can be selecting clusters of information that match the task's topic or re-ordering search results based on the current information need. Differentiating is also called *distinguishing* in which the source from which the information comes is noted.

Filtering. Using some personal criteria to increase the precision of the retrieved information. An example on the web can be using particular search keywords or restricting the search to a particular type of documents or a particular date.

Extracting. Extracting is the process of identifying specific information of interest on a located source. Extracting can be something like saving a web page or printing parts of it during the information seeking process. Copying information from a source such as an article and pasting the information in a different document can also be considered extracting.

Verifying. On the web, everyone can post just about anything. Consequently, verifying is most needed during the process of information seeking. An example of web source verification can be by extracting keywords from a source such as a web page and searching for information that confirms the source's trustworthiness on another web page. Verifying is similar to filtering where the precision and relevancy of information is targeted. An example of filtering is by restricting search to certain time or file type while using a search engine.

Ending. Ending defines the end of the information seeking process. Creating summaries and concluding notes can be signs of ending.

2.1.2.Marchionini's Model

Marchionini (1995) stated that the process of information seeking consists of several activities (sub-processes) that start with the recognition and acceptance of the problem and continues until the problem is resolved or abandoned. Figure 2 shows those activities. The sequence of the sub-processes can be the default trend; however, each sub-process may happen at any time, be active at all times, be temporarily frozen, or may call other sub-processes. The following is a description of each sub-process.

Recognize and accept an information problem. The problem can be the need for information or demand for resources. Recognizing the need for information is user-centered and can be either accepted or suppressed. Accepting the problem is the first step for deciding to search for information. Suppressing it is delaying, ignoring, or postponing the search. Accepting the problem leads to understanding it.

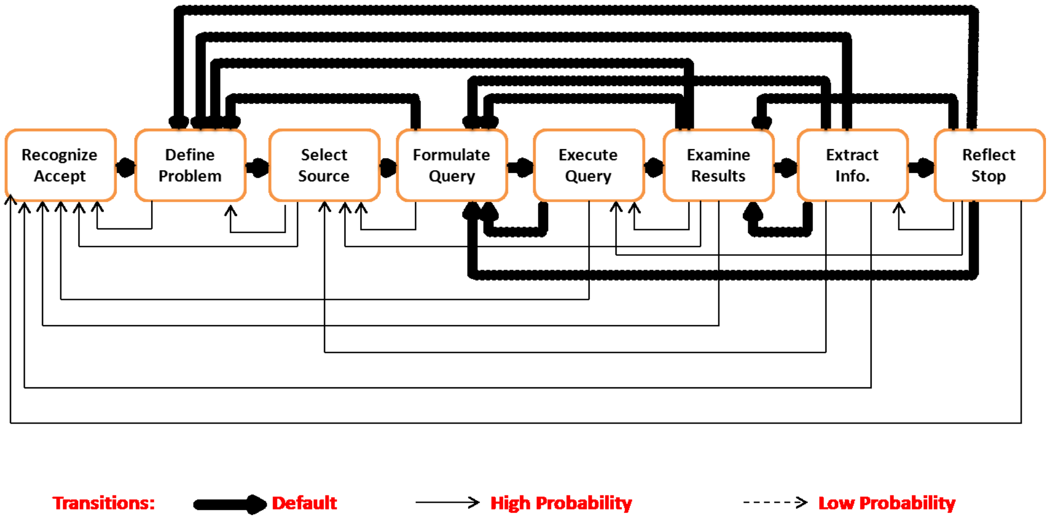


Figure 2. Machionini's information seeking model.

Define and understand the problem. This activity is very crucial during the process of information seeking. Understanding the problem depends on the user expertise of the task and the system. This sub-process can be a major cause to user satisfaction or frustration (Doll and Torkzadeh, 1988). An information seeking problem can be defined by identifying related knowledge or similar problems. By defining and understanding the information seeking problem, the information seeker may create an expectation of what the information looks like or what the answer will be which influences the selection of the search system in the next step.

Choose a search source. Selecting a search source depends on the user's expertise of the task domain and the answer expectations that may have been initiated during the previous activity. With regard to the user's expertise, the study conducted in (Alhenshiri, et al., 2011) showed that more experienced users tend to user different search engines for different tasks on the web. For example, Bing was shown to be used for searching for images. General web users tend to turn to a default search engine for all kinds of tasks (Alhenshiri, et al., 2011). This finding complies with Machronini's model description. Information seekers may also consult several search systems as they continue the process of solving their problem.

Formulate a query. Formulating a query is conveying the seeker's information need to the system in the form of keywords; a process in which two kinds of mapping are involved: a) semantic mapping of the users need represented by their selection of keywords to the system’s vocabulary that represents the information content; and b) action mapping of the user’s strategies to the features of the system’s interface. Semantic mapping is preceded by taking the user’s understanding of the problem, creating the user’s need, and transforming that need into the form of keywords (query) that can be mapped to the system’s content.

Several advances have been achieved on the web such as dynamic suggestions of query terms, multiple query submissions, and visualized query reformulation. This kind of mapping is similar to answering the question what content the system has and which matches the content of the query? Action mapping, however, is similar to answering the question how; i. e. how can the user use the system’s input to take the information need to features available by the system’s interface? Both mappings in the query formulation stage depend on the user’s expertise. More professional users may have more terms to use and may be more able to interact with the system by reformulating queries. A difficult mapping problem is the mapping of the task (not the information need) vocabulary to the system’s vocabulary.

Execute query. Executing search is the sub-process following formulating a query. Search execution aims at achieving the goal of the information seeker’s task (completely or partially followed by query re-formulations). On the web, this activity is performed by the user in the hope of getting results relevant to the information need on the way towards solving the problem initially understood and accepted by the user.

Examine results. As a result of search execution, the user is presented with one or more search results. On the web, search results are called search hits. Examining the results depends on the user’s expectation of the results, the degree to which the results solve the accepted problem, and the level to which the task’s goal is achieved through the search process. The number of hits, their type, and their relevance to the task at hand play a substantial role in the examination process. The user’s reaction to the results may lead to one of several directions as shown in Figure 2. On the web, substantial progress in supporting results examination has been made. The concepts of ranking, using multiple results features, visualization, and clustering in addition to other techniques have made examining the results much more efficient, effective, and satisfactory.

Extract Information. After examining the results of a search activity, the relevance of the document to the task and the level to which it satisfies the task goal determine the amount of information to be extracted from the document for the task. On the web, reading, scanning, cutting, and saving are examples of activities involved in the extraction process. After extraction, the information is integrated in the seeker's knowledge domain. The following step may involve further rescanning of the same document, re-examination of the results, execution of another search process, or stopping and accepting that the problem is solved.

Reflect, iterate, or stop. It has been shown that when the concept of information gathering is involved in the seeking process, the task is rarely completed through one search process (Mackay and Watters, 2008, Alhenshiri, et al., 2011, 2012b). Amin (2009) showed that information gathering is a heavily search-reliant task that may involve several iterations of search activities. Generally, the initial set of results provides feedback usually for further formulations and query iterations to the search system. Navigation may become a part of the search progress on the web (Alhenshiri, et al., 2010d, Manning, et al., 2009). Stopping the process of information seeking is connected to achieving the goal of the task or abandoning the information seeking process.

2.1.3.Wilson's Model

Wilson and Walsh (1996) model of information behaviour differs from other models by suggesting more high-level information seeking search processes: passive attention, passive search, active search, and ongoing search. These search processes are illustrated in the following discussion.

Passive attention. This information seeking activity may occur when the user does not intend to locate or seek information; yet the information become available and information acquisition takes place. An example could be listening to radio or watching the television. An example of this kind of activity on the web could be serendipitous search.

Passive search. The kind of search that is intended for locating or seeking one piece of information; nonetheless, information relevant to the user and which is not originally intended is located. An example of such activity can happen on the web such as in the case of serendipitous search.

Active search. This is the most common type of search activity. A user seeks information on active basis while having the intention to find such information. This activity is common on the web. However, the model at hand did not consider providing examples of active search that may take place on the web. Searching the web using a search engine or a search service on a web site are examples of active search.

Ongoing search. While active search establish the search activity for seeking information; ongoing search happens for updating what was established through active search. An example from the web environment would be establishing active search to find a web page and saving it in the bookmarks. Monitoring the page for reflecting updates is ongoing search for seeking newer information on that page. However, the model did not consider the case of the web. In the case of the web, ongoing search may include navigating through a hierarchy of pages, continuous submission of search queries and so forth.

2.1.4.Wilson's Combined Model

Ellis (1993) and Kuhlthau (1991) provided stages as additional parts of the information seeking process. As a result, Wilson (1999) combined their work and suggested different varying sequences in the information seeking behaviour as shown in Figure 3.

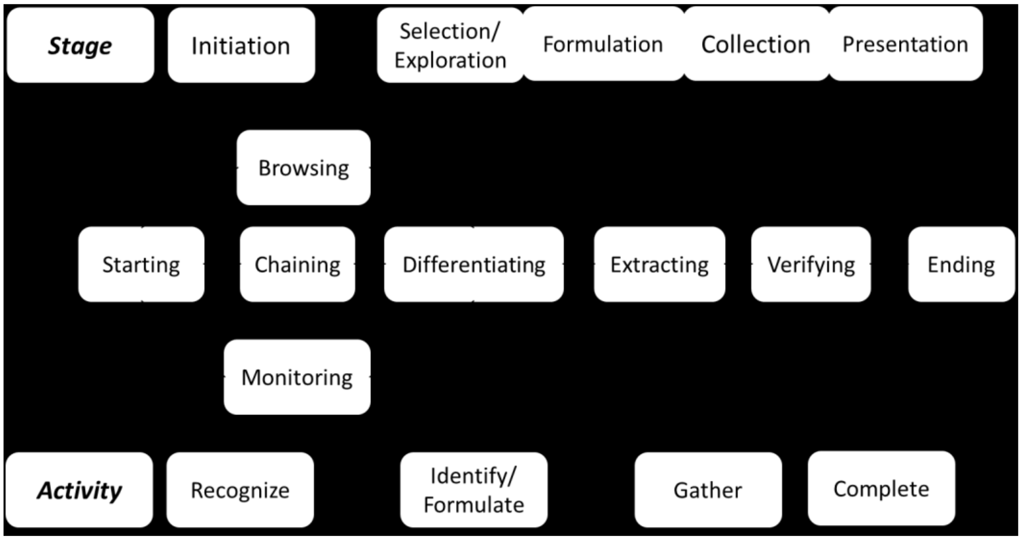


Figure 3. Wilson's combined model.

Although these models provided good characterizations of users' information seeking activities, several activities that users perform on the web are not considered in the description of those models. The variations of these models and the ongoing modifications make it hard to choose an appropriate characterization. To understand and model the different activities users perform specifically on the web while seeking information, several other frameworks have been suggested (). The following section explains some of the user task-specific models for the case of the web.

2.2. MODELS OF USER TASKS ON THE WEB

At the beginning of the 2000s, research (Sellen, 2002; Broder, 2002) started looking at a concept that exceeded a single activity of web search to a complete task. Researchers have examined user web sessions, user behaviour, the effect of work tasks on information search behaviour, user goals, general web activities, search activities in particular, and ultimately high level tasks in order to provide practical design recommendations for web tools.

The effect of work task on the interactive search behaviour of web users was investigated in the works of Yuelin, et al. (2008) and Liu, et al.(2010). Yuelin, et al. (2008) chose simulations of real work tasks for the experiment due to the overwhelming number of possible work tasks in the real world. Work tasks were categorized into six facets based on the principle that tasks vary with respect to the product and objective complexity. Product has three values: physical, intellectual, and decision/solution. The value of physical was left out since the subjects of the experiment were university students. Objective has three values: high complexity, moderate complexity, and low complexity. Six tasks were used. The experiment evaluated the effects of work tasks on users' general interactions including general uses of the information retrieval system and resources, result pages viewed, items viewed, and items selected for the task. In addition, the experiment evaluated the effect of the work tasks on users' interactions with the web resources, library resources, and query-related interactive behaviour. Although the study showed which facets of the work task affected what aspects of the interactive information retrieval process, there are several other objectives that yet need to be examined such as what factors affect the selection of the search process, the search sources, and the organization of knowledge during the task.

He and Goker, (2000) and Jansen, et al. (2007) studied search sessions to identify boundaries among user search sessions, and to be potentially able to decide on the user search goal in each session. Both studies intended to improve the effectiveness of the search process by providing more suitable results to the user's goal. He and Goker (2000) used the logs of the search engine Altavista[1] to identify boundaries that separate search sessions by relying on time, IP address, and the number of iterations per activity. Session intervals generated form the analysis of the log records were compared to judgments by humans. The ultimate goal of the study aimed at providing automated approaches to identify a session as a set of related activities.

Jansen, et al. (2006) used query interactions from the *Dogpile*[2] meta search engine to identify the most effective approach for defining web sessions. Three methods were examined: User IP address and browser cookie; IP address, browser cookie, and time cut-off; and IR address, browser cookie, and query content change. After comparing the session boundaries identified by each of the three methods to those identified by humans through manual classification, the results showed that the third method in which the content change is employed was the most effective. The change of content in addition to the IP address can be used as effective identifiers of where a session and a web task starts and where it ends. This would lead to building information retrieval systems that may help the user in reformulating queries during web search and achieving more accurate results as indicated in the findings of the study.

In addition to identifying web sessions and determining boundaries of where web activities of a single user aiming at a certain goal start and end, research has focused on the higher level categorization of web activities. Several models and frameworks (Bystrom and Hansen, 2005) have been proposed in the literature to group and categorize user activities on the web into tasks. The goal is to lead research to more practical and effective designs of web tools intended for improving each task identified. The following is a review of taxonomies and models of web tasks.

[1] http://www.altavista.com/
[2] http://www.dogpile.com/

2.2.1.Broder’s Taxonomy

Broder (2002) studied different user interactions during web search and identified three types of tasks based on the queries submitted by users. He used a questionnaire that collected responses from 3190 users about the goals of their search activities. In addition, he used log analysis of 400 queries that were inspected manually. The research identified three different kinds of tasks: navigational, informational, and transactional. Navigational queries are submitted to reach a particular website or page the user has in mind. The user knows the site or page because they have seen it in the past or it is assumed to exist. An example of a navigational query would be something that is looking for a company’s website. A query such as ‘toshiba’ is assumed to possibly have ‘http://www.toshiba.com’ as a target. This type of queries has one ‘right’ answer and the type of search is referred to as a ‘known item’ search in classical IR. According to Broder (2002), this type of tasks represents between 20% and 24% of user tasks on the web.

Information queries are submitted by users who assume that such information is available on the web. In the case of informational tasks, it is assumed that the information is static and that the user finds it and reads it with no further interactions. Informational queries may be extremely wide such as ‘cars’ or ‘San Francisco’ or extremely narrow such as ‘Toshiba satellite laptop overheating problems’. The target of queries submitted looking for information can be a collection of documents rather than a single good document. This type of task represents between 39% and 48% of the overall tasks on the web according to Broder (2002). Transactional queries are intended to reach a website or web page where further interactions will take place. A shopping task may start with a query that aims at reaching a shopping site and continues with several interactions such as looking for different products and prices. Other examples may involve finding web services or locating download sites. According to Broder’s taxonomy, this type of task represents between 30% and 36% of user tasks on the web. The different tasks identified in Broder (2002) are shown in Figure 4.

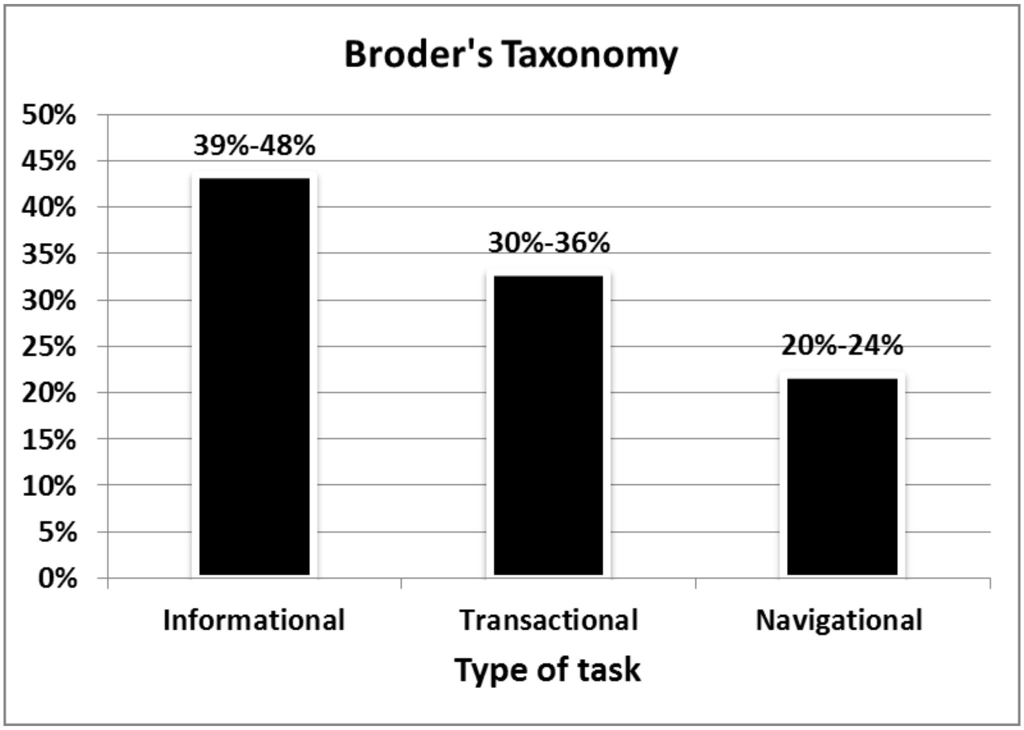


Figure 4. Broder’s taxonomy of web search tasks (2002).

2.2.2.Sellen’s Taxonomy

Sellen, et al. (2002) studied the web activities of 24 knowledge workers over two days. Participants were asked to describe their web activities. The classification resulted in six main categories: finding, information gathering, browsing, performing a transaction, communicating, and housekeeping. The study used knowledge workers who were interviewed and questioned about their daily web activities. A knowledge worker was defined as ‘*someone whose paid work involves significant time gathering, finding, analyzing, creating, producing or achieving information.*’ Information was defined as ‘*anything from documents, policies, plans, and presentations to drawings, designs and graphics.*’ The study concluded on six activities those workers perform on the web which are described as follows:

Finding: finding is seeking to locate a specific piece of information (a fact) such as phone number or a product name. This type of task represented 24% of the total web tasks performed by knowledge workers in the study.

Information Gathering: gathering is ‘finding’ but less specific information than the case of *Finding* described above. Gathering information can be for comparing, choosing, or deciding about a topic. An example can be gathering material for writing a document or preparing for a meeting. This type of task accounted for the largest portion of the tasks (35%).

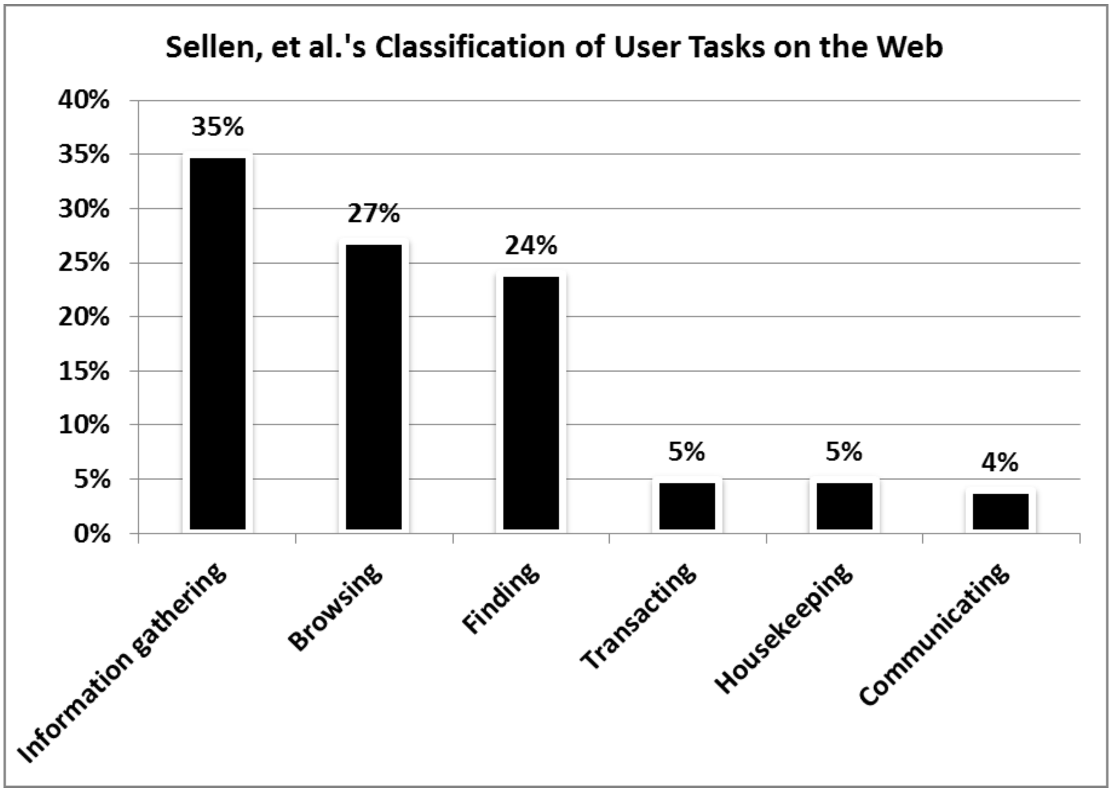


Figure 5. Sellen's taxonomy of user tasks on the web (2002).

Browsing: When going to a page or a site only to be informed, stay up to date, or be entertained, the type of task is called browsing. A user browsing the web usually does not have a specific goal in mind. Examples of browsing tasks can be navigating through a newspaper or following an interesting link. In the study, 27% of the tasks inspected in the study were browsing tasks.

Transacting: A transaction is a task in which the user uses the web to make a bank transfer, pay a bill, order a product, or download software. This task accounted for 5% of the overall tasks in the study.

Communicating: using the web for chatting, conferencing, or being in a discussion group is considered communicating. This task represented 4% of the study tasks.

Housekeeping: This task involves regular activities such as making sure that links are working properly, that certain sites are up to date, or checking an email message. This task accounted for 5% of the overall tasks.

Although most of the tasks performed by the knowledge workers in the experiment fell into one of the six categories discussed above, some tasks belonged to two or more types. For example, locating a product is first considered a finding task; then, it was also considered a *transaction* when the user bought the product. The taxonomy of tasks identified by Sellen, et al. (2002) is shown in Figure 5.

2.2.3.Rose and Levinson's Classification

Rose and Levinson (2004) identified a framework for user search goals using ontologies in order to understand how users interact with the web. A sample of queries (three sets each containing 500 US English queries) from the search engine Altavista logs was taken and analyzed for creating a preliminary set of goals resulting in a goal framework. The framework was further revised and categories of goals were either modified or added by analyzing further queries. The results of the revisions indicated that the goals of the search queries fell into a hierarchical structure of which the top level resembled Broder's (2002) Taxonomy of web tasks. The aim of the analysis was to examine to what extent the percentages of web activities that belong to each type of task has changed over time. The goal framework is described in Table 1.

Three types of tasks were identified in the query logs used by Rose and Levinson (2004). They defined a task with the goal of navigation as the task in which the user seeks a page of an institution or organization. The search query must be intended to find a website the user has in mind which indicates that queries underneath tasks of navigational nature must have the name of the organization in question. A task was considered informational if the goal was to obtain information about a topic. This type of task involves answering questions of both open and close-ended nature such as asking for advice or willing to learn about a particular topic. Indirect informational tasks may involve queries of the type 'find out about'. '*The desire in this type of task is to locate something in the real world or simply to get a list of suggestions for further research*' (Rose and Levinson, 2004). This type of task represented 61.25% in the experiment conducted by Rose and Levinson

(2004). Finally, if the task intended to find something other than information, it is considered for resources. This type of task involves queries looking to download certain material or to obtain something such as a recipe or song lyrics. The framework developed in this experiment lead to associating goals with queries using the concept of ontologies. The main task types are shown in Figure 6.

Table 1. The search goal hierarchy. Queries are only assigned to leaf nodes. All examples are taken from actual AltaVista queries.

SEARCH GOAL	***DESCRIPTION***	***EXAMPLES***
1. Navigational	My goal is to go to specific known website that I already have in mind. The only reason I'm searching is that it's more convenient than typing the URL, or perhaps I don't know the URL.	aloha airlines duke university hospital kelly blue book
2. Informational	My goal is to learn something by reading or viewing web pages	
2.1 Directed	I want to learn something in particular about my topic	what is a supercharger
2.1.1 Closed	I want to get an answer to a question that has a single, unambiguous answer.	2004 election dates
2.1.2 Open	I want to get an answer to an open-ended question, or one with unconstrained depth.	baseball death and injury why are metals shiny
2.2 Undirected	I want to learn anything/everything about my topic. A query for topic X might be interpreted as "tell me about X."	color blindness
2.3 Advice	I want to get advice, ideas, suggestions, or instructions.	help quitting smoking walking with weights
2.4 Locate	My goal is to find out whether/where some real world service or product can be obtained	pella windows phone card
2.5 List	My goal is to get a list of plausible suggested web sites (i.e. the search result list itself), each of which might be candidates for helping me achieve some underlying, unspecified goal	travel amsterdam universities florida newspapers
3. Resource	My goal is to obtain a resource (not information) available on web pages	
3.1 Download	My goal is to download a resource that must be on my computer or other device to be useful	kazaa lite mame roms
3.2 Entertainment	My goal is to be entertained simply by viewing items available on the result page	xxx porno movie free live camera in l.a.
3.3 Interact	My goal is to interact with a resource using another program/service available on the web site I find	weather measure converter
3.4 Obtain	My goal is to obtain a resource that does not require a computer to use. I may print it out, but I can also just look at it on the screen. I'm not obtaining it to learn some information, but because I want to use the resource itself.	free jack o lantern patterns ellis island lesson plans house document no. 587

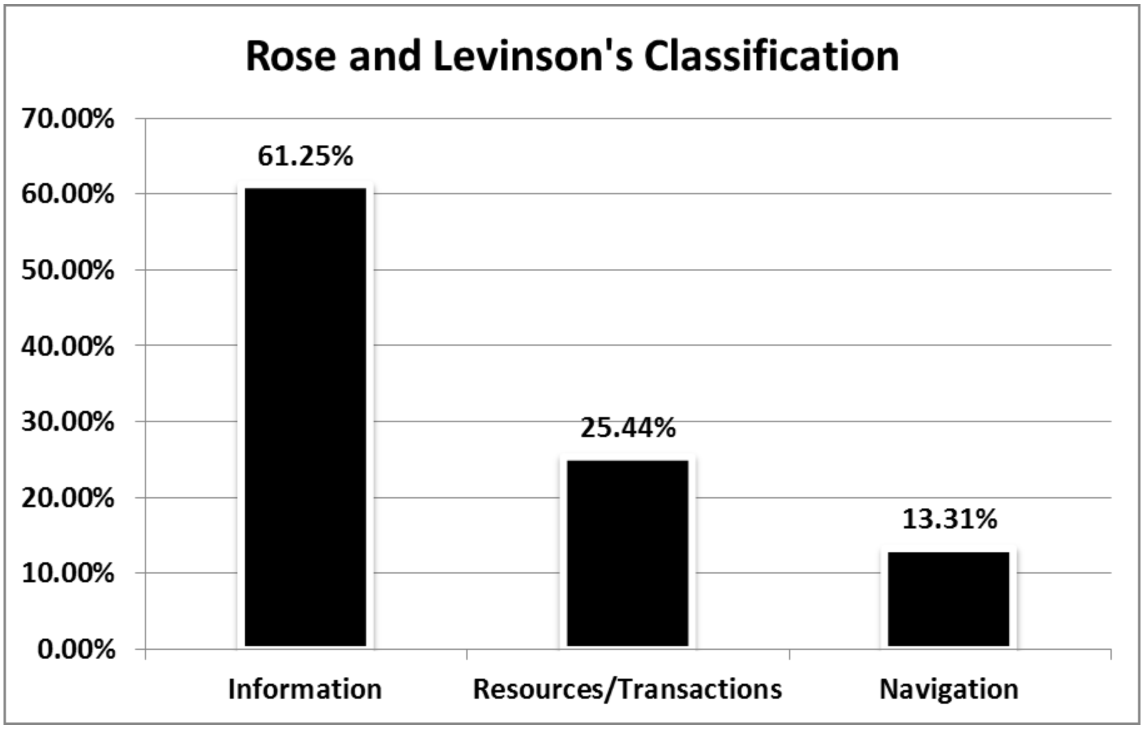


Figure 6. Rose and Levinson's classification of tasks (2004).

2.2.4.Kellar's Framework

Kellar, et al. (2007) investigated user activities on the web to develop a task framework. The research identified possible categories of user tasks using a pilot study and a focus group prior to a field study. Five categories were identified: fact finding, information gathering, browsing, performing a transaction, and an unidentifiable type of tasks tagged with 'other'. A field study was conducted in which participants were asked to describe their usual web activities using the categories of tasks concluded in the focus group. Every user provided a description of the tasks they performed and the interaction data was logged for further analysis. The percentages of tasks on the web as a result of the study are shown in Figure 7.

Fact Finding is the kind of task in which the intention is to find a factual piece of information such as the weather conditions, song lyrics, or course material. This type of task is usually repeated (55.50% repeated at least once), and it was described by participants using terms such as 'checking', 'finding', or 'looking for'.

Information Gathering is a task in which the user intends to find information about a topic such as researching a new purchase or locating information about a course or a project. This type of task was shown to be frequently repeated and to take longer durations than other tasks.

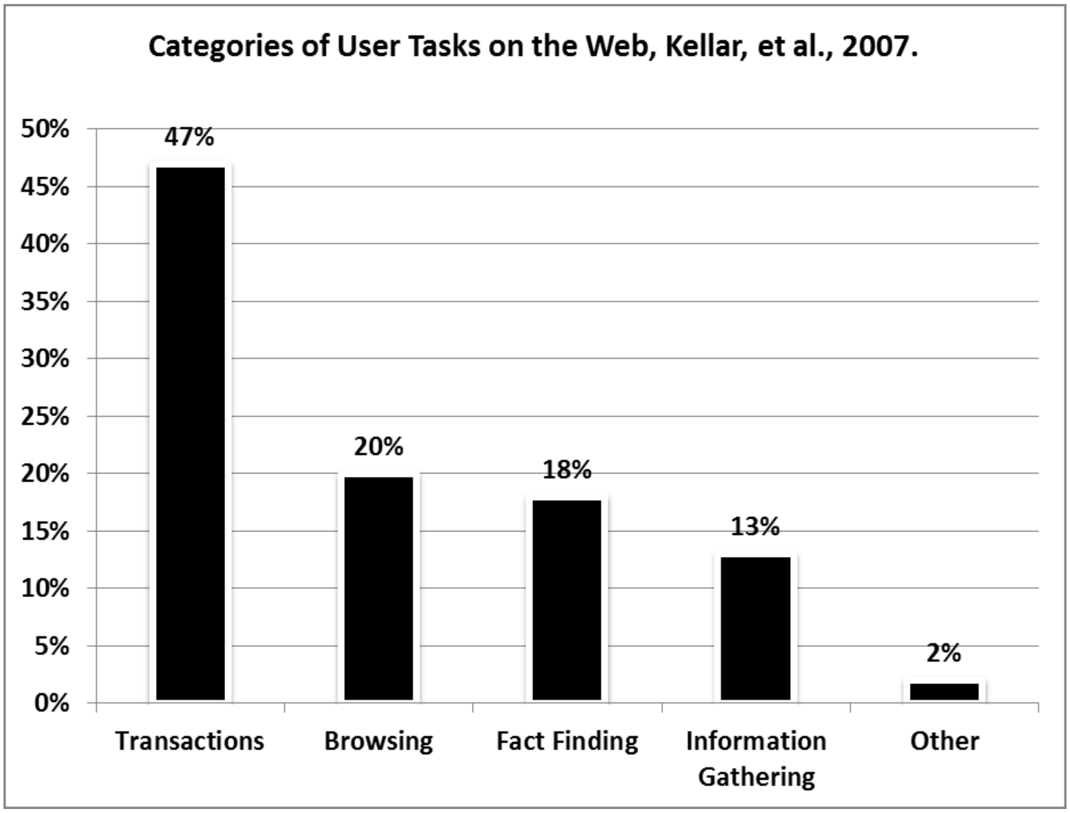


Figure 7. Kellar, et al.'s categorization of user tasks on the web (2007).

Browsing is the highest repetitive task (84.4%) and it is mainly concerned with navigation of habit. Participants in Kellar, et al.'s (2007) experiment used terms such as 'looking for' and 'reading' during this type of task. They usually navigated in habitual sequences such as when reading the news. Hobby and travel related interests were commonly related to the task of browsing.

Transaction tasks involved emailing, downloading, and online bill payments. This type of task accounted for 46.7% of all web usage. Transactions were highly repeated tasks.

Other, is a category that contains tasks that do not belong to any of the above categories. An example of this type of task is viewing a web page during development.

There are several classifications, frameworks, and taxonomies of tasks users perform on the web. Those classifications seem to disagree on certain types of activities and the labeling of the tasks identified. A possible cause of the disagreement is the changes to the user behaviour and the tools associated with the execution of tasks over time. The evolution of the web and the emergence of new genres may require changing the way the user completes their tasks on the web. It may also cause the emergence of new types of tasks and differences in the characteristics of current tasks. After identifying the high level user tasks on the web, exploring each type of task further is required to improve the effectiveness of tools used for completing those tasks. In particular, the task of information gathering—for the reasons discussed in Chapter 1—was selected in this research to be further studied and investigated.

2.3. A Review of the Task of Information Gathering on the Web

There has been no clear definition of the task of information gathering. The current definition is a simple description of the task that differentiates information gathering from other types of tasks. Information gathering tasks are known to involve collecting information possibly of different types from different sources to achieve an overall goal. Information gathering tasks are mostly search-based as shown by Kellar, et al. (2006) and Amin (2009). Information gathering is recognized as the most frequent task in re-finding information on the web (Kellar, et al., 2006). This type of task was revealed to be the goal of the search process in 61.25% of the time according to Rose and Levinson (2004).

Information gathering tasks have been studied as a part of user interactions with the web for searching and navigation as discussed by Kules, et al. (2008) and Alhenshiri, et al. (2010f). Research has also investigated some general aspects of the information gathering task. For example, Yamada and Kawano (2009) used sections in web pages located for an information gathering task to extract links to other pages. The target pages are considered a part of the user plan for the task and suggested to the user to continue gathering related information. In a similar approach, Bagchi and Lahoti (2009) used hyperlink connectivity among web pages to assist users in gathering information on the web. They argued that providing links to pages currently being viewed by the user can facilitate the process of information gathering. However, the only subtask of information gathering considered in those two studies was locating web information, i.e. finding.

Dearman, et al. (2008) investigated the subtask of finding sources of information during information gathering tasks. Re-finding information on the web was also investigated either with respect to locating previously found results (Tauscher and Greenberg, 1997; Tyler and Teevan, 2010; Badesh and Blustein, 2012) or monitoring web sources of information (Kellar, et al., 2007). Issues with how users deal with information gathering and how they manage their time for the task were discussed in the work of Murphy (2003). Addressing the problems of information mismatching and overloading during information gathering using concept-based personalized techniques was discussed in the work of Tao and Li (2009). They suggested that improvements are needed for the representation and acquisition of user profiles in personalized web information gathering. Finally, decision making was investigated and considered by Zilberstein and Lesser (1996) as an intermediate step in information gathering tasks.

With respect to research regarding how users gather information on the web, several questions remain open for further investigations. The concept of information gathering remains unclear with regard to the effectiveness of tools used for gathering and comparing web information and the challenges the user encounters during the gathering process. Moreover, the definition and components (subtasks) of the task of information gathering remain unclear. Most of the research conducted in web information retrieval attempted to improve aspects of the subtasks under information gathering without considering the contribution of the context of the whole task to the gathering process.

Prior to discussing a formal definition of the task of information gathering that is developed during this research; the following section explores investigations that used visualization and clustering in examining aspects of the information gathering task on the web. The discussion will also cover two important topics in information gathering, namely re-finding information and organizing information for the task. The discussion covers work related to the studies conducted in this research.

2.4. Visualizing, Clustering, Re-finding, and Organizing Web Information

In the context of web information gathering, there has been no specific focus in the literature on the effects of visualization, clustering, and the concepts of re-finding and organizing information on the effectiveness of users performing web tasks for information gathering. However, visualization and clustering have been investigated for improving the effectiveness of web search techniques . In addition, re-finding is a factor that has been studied on its own, and there are several techniques intended for improving re-finding of information on the web either locally on the web browser or on the entire web through the use of search engines. Information management and organization is an important topic that has been less considered in investigations related to web information gathering tasks. The following sections illustrate the use of visualization and clustering in web information retrieval in addition to re-finding and information management techniques used for accomplishing different tasks.

2.4.1.Visualizing Web Information

Visualization is a concept that has been a focus of research in information retrieval (Card, et al., 1999; Roberts, et al., 2002; Sugiyama et al., 2004; Cai, et al., 2004; Nguyen and Zhang, 2006; Friendly, 2008). Information visualization is suggested to improve users' performance by harnessing their innate abilities for perceiving, identifying, exploring, and understanding large volumes of data (Card, et al., 1999; Friendly, 2008; Alhenshiri and Blustein, 2010a, 2010b). There are several prototypes that have been investigated for improving the effectiveness of web search results (Kules, et al., 2008). Teevan, et al. (2009) investigated the use of *Visual Snippets* in web search results presentation and compared the effectiveness of this approach to the conventional text snippets provided by search engines such as Google. The results showed that combining text with the most important images on a web page may help users recognize the page more easily and be able to select pages of interest more effectively. The use of a *3D City Metaphor* in the work of Bonnel, et al. (2006) also showed that users favored the visual presentation of clustered results. Visual thumbnails (snapshots of web pages) that accompany textual presentations were also shown to be effective in searching the web for revisiting (Taucher and Greenbers, 1997; Woodruff, et al., 2001, 2002).

To further reveal the relationships (similarities) among web documents to users for more effective exploration of web search results, Zaina and Baranauskas (2005) designed a visual interface (called *ReVel*) for exploring web search results. The interface used a graphical representation of web search results. Result hits are connected via links representing similarities among documents. In addition, ReVel allows users to integrate results of multiple sessions for further exploration. Though, the visualization used content similarity as the only feature conveyed to the user. Moreover, the display suffered from clutter due to visualizing similarities among all documents using graph edges.

To provide effective topical overviews of search results, Paulovich, et al., (2008) designed a search interface that supported interpretation of collections of web results. *PEx-Web* permitted the user to avoid excessive visiting to unwanted results and to discover relevant documents based on visual topic representations through visual clustering. These two approaches (Zaina and Baranauskas, 2005; and Paulovich, et al., 2008) were shown to be effective when compared to raw presentations of web search results. Finding related information to the sources of interest during information gathering tasks would benefit from these types of visualizations.

Visualization was also shown to be effective in presenting extensive numbers of results on the display using *Periscope*, a prototype investigated by Wiza, et al. (2004). The use of visualization in results presentation was also investigated and shown to be effective in search results exploration in the works of Bonnel, et al., (2005 and 2006), Zaina and Baranauskas (2005), Joho and Jose (2006), and Paulovich, et al., (2008). In addition to results presentation, visualization has been investigated in query construction and reformulation (Kawano, 2000; Havre, et al., 2001; Dörk, et al., 2009). Research showed that using visualized query terms and phrases permits users to construct more effective queries that lead to more accurate results. In the work of Kawano (2000), users were able to effectively utilize 45.5% of the terms provided visually on the display to construct alternate queries, and 84.8% of those queries consisted of more than two keywords. In addition, Grewal, et al., (2000) showed that visualizing the process of assigning degrees of significance to query terms resulted in ranking search results more accurately.

Visualization plays a role in how users explore web documents in techniques and prototypes that use visualized clusters (Kules, et al., 2008; Carpineto, et al., 2009). There are several search tools on the web that use visualization such as the search engines Gceel[3], Nexplore[4], and Viewz[5]. Visualization of web search results has been investigated in several layouts including the use of hyperbolic trees (Rivadeneira and Bederson,

[3] www.Gceel.com
[4] http://www.nexplore.com/
[5] http://www.viewzi.com/

2003), Scatterplots (Shneidermann, et al., 1996), Self-Organizing Maps (Au, et al., 2000), and thematic maps such as in the 'formally visual' search engine Kartoo[6]. Moreover, visualization has been investigated in exploring the hierarchy of web sites for search and navigation (Bederson, et al., 1996; Karim, et al., 2009; Alhenshiri, et al., 2010a). Alhenshiri, et al., (2010a) showed that providing users with the ability to see their search path while navigating to locate information of interest resulted in more effective search compared to the use of the web browser to search for information by navigation.

Shneiderman's visualization principles (Shneiderman, 1980): "overview first, zoom and filter, then details on demand" have guided several research works intended for improving web search by integrating visualized clustering. For example, Tilsner, et al. (2009) designed a search interface that provided visualized clusters of web search results while allowing users to expand the cluster overviews to see more of the search results in every cluster. They stated that results provided to unclear or ill-defined queries should be categorized in different clusters while allowing documents that belong to different topics to be placed in more than one cluster. Moreover, Di Giacomo, et al. (2008) designed a graph-based visualization of web search results that helped users with exploring topical clusters of web search results. They argued that visual analysis would allow for better exploration of search results than directory tree paradigm-based clustering that is implemented in several web search engines such as Clusty (Yippi[7]) and *iBoogie*[8]. The *radial* and *treemap* layouts were compared to a previous version of the interface that used an orthogonal layout and were shown to provide more effective exploration of web search results. All in all, finding information sources is a subtask in the web information gathering task that would benefit from such techniques; however, further investigations are needed with regard to information gathering on the web. Most of the evaluation studies that have been conducted, such as in the works of Tilsner, et al. (2009) and Di Giacomo, et al. (2008) lacked the context of a task and relied on relatively small usability tests and case studies using simple web queries.

[6] www.Kartoo.com
[7] http://search.yippy.com/
[8] www.iboogie.com

The use of visualization may have some drawbacks. The work of Alhenshiri, et al. (2010b) showed that some users complained about issues of clutter in the visual search engine. In addition, 3D visualizations can inhibit users and make interfaces more confusing (Sutcliffe and Patel, 1996; Risden, et al., 2000). Visualization can also make the exploration of search results more frustrating in case no meaningful axes are defined on the display (Kules, et al., 2008). Some visualization layouts can be unproductive such as the use of *Data Mountains* for browsing tasks as demonstrated by Cockburn and McKenzie (2002). These issues should be considered with complex tasks such as information gathering.

2.4.2.Clustering Web Information

Clustering is intended for grouping together items that share similar characteristics and attributes. In web information retrieval, clustering is meant for grouping similar documents (Manning, et al., 2008). The use of clustering has been widely investigated in web information retrieval (Katifori, et al., 2007; Ferragina and Gulli, 2005). Clustering is usually intended to provide an overview of categories in the result set. Hence, efficient subtopic retrieval is anticipated with the use of clustering in web search results presentations (Carpineto, et al., 2009). When more than one topic is desired while gathering information on the web, clustering may provide effective topic exploration in the high-level views of the result hits. Clustering can also decrease the need for scrolling over multiple pages of results and also motivate users to look beyond the first few hits (Spink, et al., 2001; K¨aki, 2005). Moreover, clustering has other benefits such as capturing meaningful themes in the search results, scalability, and domain independence (Efron, et al., 2005). All of these advantages may benefit information gathering on the web where more than one source of information and more than one type of data are required to be explored and collected for satisfying the task requirements.

In web information retrieval, clustering has been investigated in several prototypes such as in the work of Zamir and Etzioni (1999), Feng, et al. (2006), and Alhenshiri, et al. (2010f). Clustering has also been implemented in conventional search engines such as Clusty, Gceel, Northern Light[9], and Google (in the "*see similar*" feature and *Google*

[9] http://www.nlsearch.com

Wonder Wheel). Although the performance of users with row presentations of web search results is comparable to their performance with clustering-based presentations, user preference usually comes in favor of clustering-based approaches (Carpineto, et al., 2009). Interstingly, there are indications that clustering can even be more effective. Turetken and Sharda (2005) used a graph-based visualization that shows relationships between clusters. Their technique was shown to be more effective than ranked textual lists of results. Furthermore, Jing et al (2006) showed that clustering was very effective in image search. With the variety of information that is gathered on the web, the need is to investigate the role of clustering in several subtasks underlying web information gathering. Clustering can be further investigated in locating sources of information, locating and gathering further related information to the sources found, and comparing information for decision making during the task.

Furthermore, the concept of genre-based clustering is anticipated to improve the effectiveness factor in web search. A genre is a class of documents that are similar with respect to content, structure (form), and functionality (Dong, et al., 2008). Classifying documents by genre has been shown to be considerably accurate (Mason, et al., 2009). Features based on which web page genres are identified have also been studied and investigated (Ferizis and Bailey, 2006; Stubbe, et al., 2007; Levering, et al., 2008; Santini and Sharoff, 2009). Although finely grained sets of web page genres may be difficult to produce and may also differ—if produced—due to the evolutionary nature of the web, there are certain types of genres that are agreed upon in the literature (Santini, 2006). However, the effect of genre-based clustering of web search results on the relevancy, effectiveness, and precision in web search has yet to be investigated.

The web implies massive numbers of documents. Therefore, speed is a concern in online clustering. Off-line clustering, on the other hand, may suffer due to the rapid changes in the web content. As a result, clustering is usually performed by meta-search engines that use the top search results provided by an underlying index-based search engine. An example can be seen in the search engine Gceel. Moreover, generating meaningful groups and effective labels is a recognized problem in clustering (Shneiderman, et al., 2000; Kules, et al., 2008). Usually, cluster labels are generated from the titles and/or summaries of search results. In the case of using the entire document, creating a meaningful label can be very difficult due to having to deal with much more text than in the case of using summaries (Manning, et al., 2008). Finally, more difficulties arise with the issue of clustering when a document belong to multiple topics and is placed in some cluster while excluded from other relevant clusters (hard clustering is used).

The use of clustering for the purpose of improving how users gather information on the web has yet to be further examined with particular subtasks. Clustering can actually help users with selecting information sources for the task. The overview provided for web documents through the use of clustering can also help with identifying the kinds of data needed in the task. Comparing information and decision making is a subtask that can benefit from the use of clustering in addition to the subtasks of finding information and finding information sources.

2.4.3.Re-finding Web Information

Re-finding is one of the main subtasks of the information gathering task that is considered throughout this research. Previous research has focused on the design of the web browser and its capabilities to assist users in re-finding information on the web. One of the most researched navigation mechanisms in web browsing is the *back button* provided in most web browsers since around 58% of all page visitations are revisits (Kaasten and Greenberg, 2001; Teevan, et al., 2009). The essential behaviour of the back button is stack-based which takes the user back to the pages that have been most recently visited (Cockburn, et al., 2002). However, not all previously visited pages are accessible via the back button. Therefore, alternate behaviours are present in the literature such as the temporal behaviour discussed in the work of Cockburn, et al. (2002). In their work, the back button maintained a complete list of pages that have been previously visited. Moreover, the back button was further enhanced by introducing most-visited lists (Mountaz, 2000), bookmarks (investigated by Kaatsen and Greenburg, 2001), and sidebars (Berkun, 1999).

The most-recently-visited-list approach provides recently viewed pages to the user as a quick list of links to which the user can return to revisit a page he or she previously visited (Tauscher and Greenberg, 1997). Most-recently-visited lists are intended to provide a quick viewable queue of pages, which also implies navigating in sequence, instead of the hidden list of the back button. Moreover, sidebar mechanisms, such as the explorer bar in the work of Berkun (1999), deal with the issue of insufficient support for helping users find and return to individual web pages. Berkun (1999) stated that users rarely make effective use of the lists of most-recently-visited pages while navigating on the web. Bookmarks, favorite lists, and most-recently-visited lists of pages provide mechanisms for remembering and revisiting sources of web information. However, as the lists of pages grow, the value of these techniques decreases (Berkun, 1999).

Bookmarking requires that the user creates a bookmark explicitly every time he or she wants to visit the same page later. Maintaining bookmarked pages becomes more difficult when the list of pages grows extensively (Yamaguchi, et al., 2004). In addition, viewing and searching lists of ordered bookmarks is difficult. Bookmark lists typically include either the actual URL or the title of the page, which may not match the mental image the user has for the page. Consequently, Kaasten and Greenberg, (2001) introduced an interesting approach that combined the behaviours of bookmarks, back button, and browsing history in one model. To enhance the user's ability to manage visited pages and bookmarks, Kaasten and Greenberg (2001) used one recency-ordered list of visited pages. Visual thumbnails of the actual pages were shown along with the page titles on the list to attempt to more closely match the user's mental image of the page. Moreover, bookmarks were emphasized with the most-recently-navigated pages highlighted and bolded. Similarly, Mountaz (2000) combined bookmarks, most-visited pages, short-term history, and a fourth set of unclassified pages in an integrated tool called *BookMap*, with a special emphasis put on most-visited pages. In this approach, the BookMap interface allowed its users to use the four types of pages more effectively for navigation.

To minimize the effort needed by the user to manage the back and forward activities in a navigation task, Moyle and Cockburn (2003) introduced a flick gesture-based back and forward technique. In their work, the user's effort is limited to moving the mouse on the browser screen instead of having to reach out to the sidebar on which the back button resides. Moyle and Cockburn (2003) recognized that the traditional back button suffers from the distance and targeting issues of Fitts' Law (Fitts, 1954). In their evaluation experiment, Moyle and Cockburn (2003) found that participants navigated significantly faster with the flick gesture-based approach than they did using the traditional back button.

Research has focused on enhancing re-finding information on the web locally on the web browser. However, the re-finding strategies investigated can maintain a limited number of links. In addition, the use of those strategies is limited to pages and sites of interest during particular search sessions. Therefore, searching the web for re-finding, also known as re-searching (Teevan, 2008), has been studied for assisting users in locating results that were found interesting in previous sessions. Research shows that a great deal of web search visitations is for revisiting (58% according to Kaasten and Greenberg (2001), and 81% according to Cockburn, et al. (2003)). Consequently, Teevan (2008) designed a ranking technique that kept track of the user search sessions and merged user-relevant results of previous sessions with the results of the current session based on similarities among the search queries used in both sessions. The approach was shown to be effective for re-finding search results for reusing.

Re-finding is a common activity in web information gathering tasks accounting for 53.27% according to Mackay and Watters (2008). According to Adar, et al. (2008), re-finding a page on the web depends on the page itself, the topic being searched, and the intent of the user. Investigating re-finding in the context of web information gathering is further needed and it may reveal different findings that will assist the design of web search tools intended for this type of task. The study in Alhenshiri et al (2012a) discussed is Chapter 4 will further demonstrate the need for investigating re-finding during web information gathering, a task that requires more than one session to complete.

2.4.4.Managing and Organizing Web Information

Managing web information is concerned with how people store, organized, and re-find web information (Elsweiler and Ruthavan, 2007). Information management systems are methods by which users find, categorize, and re-find information on daily basis. Research has considered personal information management with less focus on the web. The web implies more information to be located, stored, and relocated. It also implies the need for managing by formatting, editing, and organizing the information to comply with the task requirements.

Information management has been explored in different directions. Research has focused on investigating how users manage their information for re-finding (Jones, et al., 2003; Mackay, et al., 2005; Elsweiler and Ruthven, 2007). Knoll, et al. (2009) investigated how users view and manage desktop information in general. Jones, et al. (2008) investigated important reasons behind giving up on certain personal information management tools. Strategies users follow to manage web information in order to be able to relocate and reuse information previously found are discussed in the work of Jones, et al. (2003). Their work showed that users, while gathering web information, follow different keeping strategies to re-find and compare information later. Most users gather information over multiple sessions (Spink, et al., 1996; Mackay and Watters, 2008), which indicates the need for management strategies for preserving and re-finding such information for reuse. The variety of finding, re-finding, organizing, and management strategies users follow while seeking and gathering web information (Alhenshiri, et al., 2011) can be attributed to the fact that current web tools lack important reminding, integration, and organization schemes (Cutrell, et al., 2006).

Jones, et al. (2008) found that users abandon the use of an information management tool for one or more of five closely related reasons: visibility, integration, co-adoption, scalability, and return to investment. These reasons need to be further investigated in the case of the web. The web may reveal further reasons why users use certain tools over others, why they do not use the same tools, what tools most users actually use to keep track of their gathered information, and how they maintain the consistency of information located for the task. Other questions may include what tools are actually supportive of information organization and management during information gathering, if any? Research has had little consideration to factors that would improve how web users collect, manage, compare, and organize their information for information gathering tasks. Practical recommendations for supporting the design and implementation of web information gathering tools with respect to information organization and management are needed.

In an attempt to demonstrate the significance of the subtask of information management during information gathering on the web, a prototype called *HunterGatherer* was designed by Schraefel, et al. (2002). The main goal of the system was to allow users to have more focus on the task of information gathering by limiting the effect ofs the burden of information management during the gathering process. During the gathering process, the user is likely to collect pieces of information from web pages (the subtask of finding information on located resources). This procedure requires the user to find the sources (pages) first. The user then uses other tools or applications such as text editors to copy the required information from those pages and keep it in files, emails, and so forth. These steps are activities that belong to the subtask of managing and organizing information. HunterGatherer allowed the user to do both subtasks in one tool. It permitted the user to locate pages that contain the information of interest and also copy and edit the information required by the task in the same tool.

Grayson, and Hedrick (2001) created a multilayer browser interface to tackle the issue of managing and browsing information on the web. The interface had two windows. The first window is called the *Driver Frame* in which users performed navigation activities such as clicking on links currently displayed, and search activities such as formulating and submitting a query. The second window is called the *Viewer Frame* in which users' actions are executed. The intention of the design did not take into account managing and organizing information out of the scope of the current active session. The organization feature that was taken into consideration concerned searching bookmarks and history and the ease of adding bookmarks to the collection of information gathered.

In a field study, Elsweiler and Ruthavan (2007) asked participants to describe their re-finding tasks. Tasks were either related to email or web re-finding. Three types of tasks were recorded: lookup tasks, item tasks, and multiple item tasks. Lookup re-finding is meant for locating a piece of information about which the user may or may not know. Item re-finding tasks are concerned with a single piece of information that the user knows and has encountered before. Multi-item re-finding is meant for locating multiple pieces of information that are known to the user. An interesting finding in addition to creating a model for re-finding tasks, is developing a common approach that research can utilize to create simulated tasks based on realistic ones. Moreover, Elseweiler and Ruthavan (2007) found the difficulties associated with performing a re-finding task depended on how long it had been since the information was originally found. The re-finding tasks are shown in the model illustrated in Figure 8.

According to Elsweiler and Ruthavan (2007), there are two main difficulties associated with evaluating methods intended for assisting users in the process of information gathering. First, the use of information gathered by users introduces privacy issues. Users are usually reluctant to share personal information for evaluation purposes. Second, the uniqueness of collections of information gathered by users makes creating evaluation tasks that apply to all users rather difficult. Understanding information management at the task level is a key to effective evaluation techniques (Capra and Perez-Quinones, 2006).

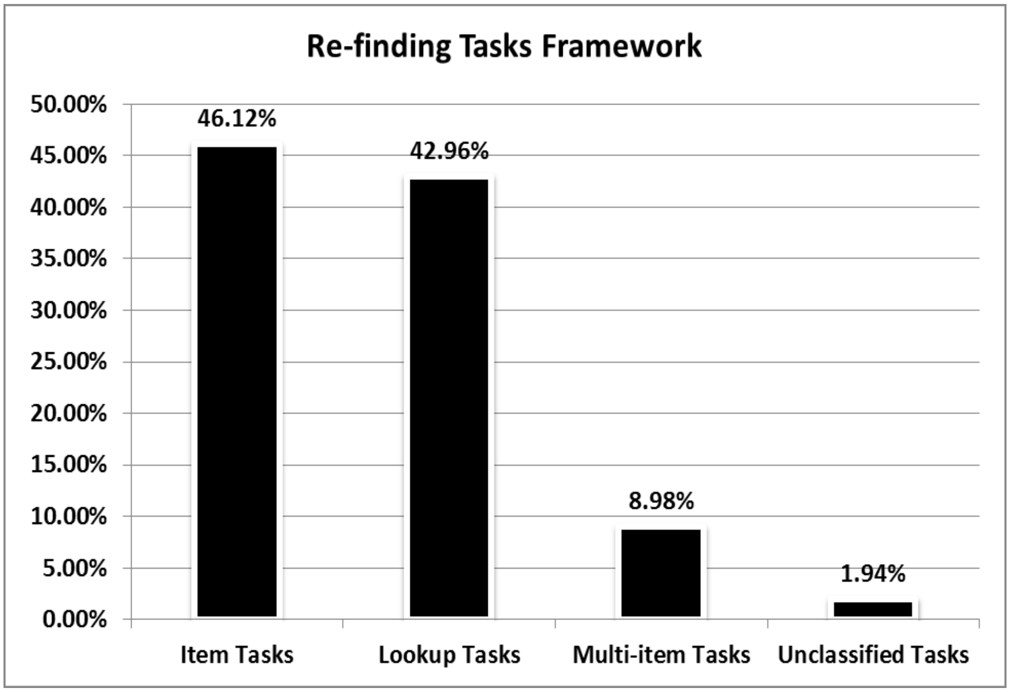


Figure 8. A framework for re-finding takes (2007).

On the web, research has only considered the case of managing and organizing information for re-finding (Jones, et al., 2003). How users organize and manage information for editing, formatting, keeping, and other gathering activities has had little consideration. Further investigations would reveal design characteristics regarding tools and features needed for improving the process of web information organization and unleash challenges users encounter with current web tools.

2.5. Summary

Research identified activities users perform to seek information. Moreover, the kinds of tasks users perform on the web have been identified and reinvestigated over time. The task of information gathering was shown to be very frequent and it requires further investigations. Studying the subtasks that comprise the overall task of information gathering may permit for better understanding of this type of task. In addition, it may allow for further improvements in the fields of information retrieval and human-computer interaction since a great portion of users' search activities on today's web are considered parts of a broader task such as information gathering.

To further understand the process of web information gathering, and to investigate possible improvements to search tools intended for this type of task, the following chapter identifies and illustrates the subtasks involved in the information gathering task. Those subtasks were developed into a framework that helped with identifying issues with each part of the task of information gathering and directed the research throughout this dissertation. A new definition of the task of information gathering on the web is provided followed by an illustration of the process of information gathering on the web.

CHAPTER 3 INFORMATION GATHERING

This chapter introduces a model of the task of information gathering. In this model, the subtasks that represent the building blocks of the task of information gathering are identified. A new definition of the task of information gathering is also developed followed by a discussion of the task process.

3.1. Subtasks in the Information Gathering Task

The information gathering task can be studied more effectively by identifying and further investigating the subtasks comprising the overall task. Based on studies conducted to investigate activities related to information gathering on the web (Kellar, et al., 2006; Mackay and Watters, 2008; Kules and Capra, 2008; Alhenshiri, et al., 2010b; Alhenshiri, et al., 2010c, Alhenshiri, et al., 2010f), the model shown in Figure 9 was initially created. The model was then further refined as discussed below.

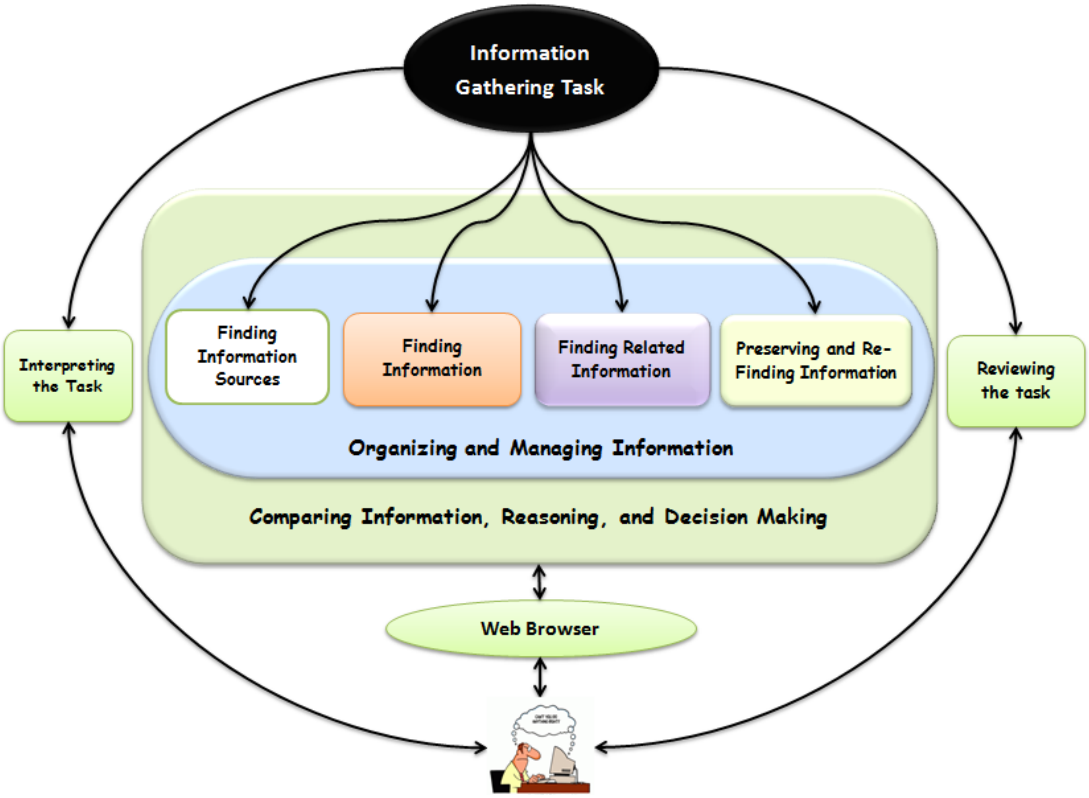

Information Gathering Task
Interpreting the Task
Finding Information Sources
Finding Information
Finding Related Information
Preserving and Re-Finding Information
Reviewing the task
Organizing and Managing Information
Comparing Information, Reasoning, and Decision Making
Web Browser

Figure 9. The initial model of the information gathering tasks.

3.1.1.Interpreting the Task

Web information gathering tasks are cognitively intensive and can be of varying degrees of complexity. To start performing an information gathering task, the user has to make a decision about the information required in the task (Bell and Ruthven, 2004), the plan desired for accomplishing the task, and the tools to be used for completing the task. Interpreting the task includes identifying information needed in the task, information about how to achieve the task and fulfill its requirements, and information about how to make a decision regarding completing the task. In addition, the user's interpretation determines the tools used in the task and their effectiveness.

3.1.2.Finding Sources of Information on the Web

The web search engine is the tool predominantly used for this subtask (Teevan, et al., 2004; Amin, 2009). Users convey their information need to the search engine in the form of a query and receive a set of resources that match the search query but not necessarily satisfy the user information need (Manning, et al., 2008, Hoeder, 2008). A study comparing user search behaviour showed that 55% of users' search behaviour involves keyword search to locate sources of information instead of typing-in a URL to the web browser (Teevan, et al., 2004). In addition, 57% of internet users use search engines daily (Hsieh-Yee, 2001; Kim, 2008). Therefore, the search engine is recognized as the tool used most for this subtask. The rest of the subtasks in information gathering are performed by the user on the web browser using different features in addition to the use of other applications.

With regard to finding sources of information, research has focused on improving the relevancy of web search results to match the user information need (Manning, et al., 2008). There are several aspects of the web search process that have been investigated including indexing (Srihari, et al., 2000), query matching (Spink, et al., 2001 ; Kawano, 2000), search results ranking (Zhuang and Cucerzan, 2006; Zitouni, et al., 2008; Wang, et al., 2009), and search results presentation (Bonnel, 2006; Teevan, et al., 2009; Alhenshiri, et al., 2010f). The latter process is concerned with interacting with the user to find sources of information required in the task. Consequently, the effectiveness in finding the intended information sources is usually concerned with how the results are presented to the user. Clustering and visualization aspects have been investigated to seek improvement to how users perceive or locate sources of information on the web.

3.1.3.Finding Information on the Web

The result hits provided by the search engine represent sources of information possibly of interest to the user. The following subtask in information gathering is locating relevant information among such sources. This stage of information gathering has been researched in several directions. On the web browser side of the subtask, results presentation has been rigorously investigated for providing recommendations for effective search interfaces. Different forms of textual presentations (Alonso and Baeza-Yates, 2003), visual presentations (Mukherjea and Hara, 1999, Bonnel, et al., 2005 and 2006), and a mix of both textual and visual presentations (Kunz and Botsch, 2002; Rivadeneira and Bederson, 2003; Suvanaphen and Roberts, 2004) have been investigated. Clustering of search results according to different criteria has also been considered (Carpineto, et al., 2009).

This subtask is usually studied as a part of the previous subtask in which there is no obvious separation between locating an information source and locating information of interest on that source. The separation is actually clear since users usually cannot make a decision just by relying on the set of hits provided by the search engine as shown in Alhenshiri, et al. (2011). In addition, a study by Alhenshiri, et al. (2010b) showed that the interface played a significant role in how users made decisions about the sources selected for the task because of how much information they were able to recognize regarding each source and the way information sources were presented. When using visualized results with several features and attributes, participants made faster decisions about their choices of information without having to open as many web pages to see the document content as they did with the raw textual presentation of search results (Alonso and Baeza-Yates, 2003). Finding sources of information is actually a different subtask from finding information because of trust and familiarity issues with web sources.

3.1.4.Finding Related Information

Finding related information to the information already identified in the sources provided by search engines is a subtask that is common during information gathering. The user finds a source of information and continues looking for task-related information in one of two ways. First, when clustering is involved in the presentation of web search hits, the user may look for similar documents to the one of interest by viewing clusters with the documents belonging to the same topic. The second approach is by following anchors on the page of interest for the purpose of finding similar information (Bagchi and Lahoti, 2009; Karim, et al., 2009; Büttcher, et al. 2010; Alhenshiri, et al., 2010a). For example, Google provides clustering in the "*see similar*" feature underneath some of the result hits. The search engine Clusty performs unsupervised clustering and presents categories of topics on a sidebar. Yahoo directories are an example of supervised clustering intended for finding related information to categories of interest. Clustering on the web is a concept intended for better topical coverage of web information which may assist users in information gathering tasks. On the web browser, following anchors on a page that link to other pages can also lead to locating related information (Bederson, et al., 1996; Karim, et al., 2009; Alhenshiri, et al., 2010a).

Finding related information is a subtask that is usually intended for gathering further information and comparing already gathered information for decision making. Consequently, it can be considered a separate subtask from locating sources of information on the web. The study conducted by Alhenshiri, et al. (2010a) showed that users followed the link hierarchy on the sources of web information they located in order to make confident decisions about the task results. Locating sources of information is usually followed by looking for related information to the content of those sources. The study also showed that different interfaces achieved different effectiveness results.

Paulovich, et al., (2008) designed a search interface called *Projection Explorer Web (PEx-Web*). Users of PEx-Web were able to discover relevant documents based on visual topic representations through visual clustering on *document maps*. The usability test showed that PEx-Web was effective in highlighting related information based on topical clustering. Consequently, finding information related to web sources located for the task is an important subtask that should be further investigated in web information gathering.

Kobayashi, et al., 2006 presented an interface to aid the process of information gathering by assisting users looking for information sources and related information. The interface used visualized hierarchical clustering (based on the content of the page) to group similar pages into a two-dimensional space on one screen. The presentation layout was a hyperbolic tree that displayed related clusters grouped in one branch on the tree. The titles of the documents in each cluster were shown to the user as well as the title (label) of the entire cluster. The interface helped users by giving overviews of more results than in the case of the list-of-hits approach followed by conventional search engines such as Google. However, this interface lacked several features such as providing different clustering criteria. In addition, hyperbolic trees usually suffer from lack of context in the case of massive collections of results and high number of clusters (branches).

3.1.5.Comparing Information, Reasoning, and Decision Making

Comparing information located for the purpose of the task happens on the browser side of the gathering process. The user performs such comparisons in different ways yet mostly by reading text on web pages (Roberts, et al., 2002). The comparison process is meant for reasoning and making decisions about the types of information required in the task (Zilberstein and Lesser, 1996). In current web techniques, comparing information requires reading much text and scrolling over multiple sources of information (web pages). Visualization is suggested to help with this process by presenting multiple features of web documents in a visual manner to assist the user with making faster and more effective decisions (Wiza, et al., 2004; Nguyen and Zhang, 2006). Clustering web information by providing meaningful labels may also assist users comparing sources of information. This subtask is involved in all of the subtasks comprising the information gathering task.

Comparing information is an important subtask that has been investigated in isolation. Suvanaphen and Roberts (2004) designed a search interface that allowed users to compare sets of results rendered to multiple queries. The objective was to permit users to observe similarities and differences among the result sets, reduce the cognitive effort that would result from switching from one result set to another, and enable them to browse more effectively. Similarly, Havre, et al. (2001) introduced *Sparkler*, a technique that visualized the results of multiple queries generated as alternatives to a user query. The interface also showed the contribution of each query alternative or component to the overall relevance of documents in the result set. The usability test showed that users preferred Sparkler to the row presentation due to the ability to observe differences between the initial query and its alternatives in the result set using the visual presentation of Sparkler. However, enhancing the effectiveness of how users compare, reason, and make decisions regarding the task information requires further investigations in the context of a complete information gathering task with a defined task goal.

3.1.6.Keeping and Re-finding Information

Information gathering tasks usually happen over the course of multiple sessions (Spink, et al., 1996; Mackay and Watters, 2008). According to Sellen, et al. (2002), 40% of information gathering tasks took more than one session to complete. Therefore, some subtasks such as finding related information and comparing information located for the task may require keeping some or all of the information for later re-finding and reusing. Research regarding re-finding information on the web has investigated several techniques in the web browser including the back button, the browser history, and the list of favorites and bookmarks. Alternative methods with similar behaviour to the aforementioned techniques have been investigated including the *mouse flick gesture* for the back and front buttons (Moyle and Cockburn, 2003), the use of *Bookmaps* for visualizing the browser history and bookmarks (Mountaz, 2000), and the use of *Landmarks* for visually presenting parts of the browser history (Mackay, et al., 2005).

Preserving search results to be involved in later search activities has also been studied in the work of (Teevan, 2008). However, it remains unknown which technique is the most effective in the case of information gathering tasks. This is so because visualization studies, such as in the works of Yamaguchi, et al. (2004) and Mackay, et al. (2005), measured how effective the presentation was in permitting the user to only find documents that have been previously bookmarked. Investigating such re-finding techniques in the context of information gathering may reveal different findings due to the existence of other factors in the context of information gathering such as the task progress, the information comparisons required for the task, and the decision making process during information gathering tasks.

3.1.7.Managing and Organizing Information

Managing and organizing information during information gathering on the web is an important subtask that has had little consideration. Users gathering information on the web may have to look for information on different sources; compare information that belong to different topics; and keep information for further analysis, comparisons, and decision making. During the process of web information gathering, users adopt different approaches for finding, keeping, relocating, reusing, editing, formatting, and saving information during the task (Alhenshiri, et al., 2011, 2012b). In addition, users may use different tools (or features in one tool) to temporarily or permanently keep the task information (Alhenshiri, et al., 2011, 2012a, 2012b, 2012c). Dealing with different sources of information including sources previously located and kept and those located at the time of the task's immediate session requires much user effort to keep track of the task requirements and satisfy the task goal. How effective current web tools in the process of managing and organizing information during web information gathering tasks should be further investigated.

3.1.8.Reviewing the Task

During information gathering, reasoning and decision making may occur at any time depending on the task, the user expertise, and the tools used in the task (Adar, et al., 2008). The process of accomplishing the overall information gathering task on the web is affected by the user's short term memory, the number of sequences required in the task, and the type of information being searched. These factors necessitate that the user reviews the task to make sure the requirements are satisfied. This subtask is an important factor that has to be further investigated in the presence of other subtasks in web information gathering. Information gathering tools and how information is provided to the user to collect, compare, and make decisions about the task should be further investigated.

Research has identified several features of the task of information gathering on the web. Those features are summarized in Table 2. The table shows each feature of the task associated with studies in which the identified feature applies. As shown in Table 2, research was able to recognize more features of the task over time.

Table 2. Characteristics of the information gathering task identified in research.

Identified Characteristics	*Studies related to identifying characteristics of the information gathering task*						
	Broder, 2002	**Sellen, 2002**	**Rose & Levinson, 2004**	**Kellar, et al., 2007**	**Amin, 2009**	**Mackay &Watters, 2008**	**Alhenshiri, et al., 2012c**
Search-reliant	✓	✓	✓	✓	✓	✓	✓
Multiple information sources			✓	✓	✓	✓	✓
High-level goal			✓		✓	✓	✓
Multiple sessions				✓		✓	✓
Complex					✓	✓	✓
Exploratory (data types)					✓	✓	✓
Multiple applications						✓	✓
Multiple subtasks							✓

3.2. Defining the Task of Information Gathering

Information gathering can be defined as a complex and highly search-reliant task with a high-level goal. It consists of more than one subtask and may take more than one session to complete. Information gathering usually needs exploring multiple sources of information to discover and gather one or more kinds of information. Information gathering may also require the use of applications and tools other than the web browser. Information gathering on the web is a complex task with high cognitive intensity. It indicates uncertainty, ambiguity in information need, and need for discovery. It also suggests knowledge acquisition, comparison, exploration, or discovery. Hence, information gathering tasks can be of different levels of complexity. During information gathering, users search the web by submitting queries, typing-in a URL, or following links on the web hierarchy.

A user performing an information gathering task has a high-level goal to collect, manage, and prepare information for the task such as in the cases of writing a research article or a project report. The goal is either directed, i.e. finding answers to questions or learning about a topic; or undirected, i.e. locating advice, services, or other information. Usually, the goal of the task implies finding, comparing, reasoning, and decision making.

During the task, users perform several subtasks each of which involves activities of different kinds. Those subtasks may happen over one or more sessions. A session is a period of time during which the user works on the task. Users gathering information on the web may have to explore sources of information (sites and pages) that belong to one or more genres. They may also have to collect data of different types to satisfy the task requirements.

The different subtasks involved in information gathering, the need for exploring varied sources to collect different kinds of data, and the need for managing the information according to the task requirements may necessitate the use of different applications in addition to the web browser.

Formally, we define an information gathering task on the web as an **8-tuple (C, SE, G, ST, S, R, D, A)** where:

- ***C:*** *is the level of **C**omplexity of the task.* C_i= {simple, moderate, complex}.
- ***SE:*** *is the type of **SE**arch.* SE_i = {submitting a query, typing in a URL, following a link}.
- ***G:*** *is the **G**oal of the task.* G_i= {directed, undirected}, **directed**= {answering a question, information about a topic, … etc.}, **undirected**= {searching for advice, locating unknown service, … etc.}.
- ***ST:*** *is a set of **S**ub**T**asks.* ST_i = {st_1, st_2, ….., st_n}, where st_i = {a_1, a_2, …, a_n}, where a_i is an activity.
- ***S:*** *is the number of **S**essions required to complete the task (time).* S_i>=1.
- ***R:*** *denotes the types of sou**R**ces of information needed for the task.* R_i= {$genre_1$, $genre_2$,…, $genre_n$}.
- ***D:*** *is the **D**ata types needed in the task.* D_i= {text, hyperlink, image, table, spreadsheet… etc.}.
- ***A:*** *denotes the **A**pplications needed during the task.* P_i= {web browser, text editor, email, calculator, …, etc.}.

The features selected to represent the task in the definition involve the most important characteristics that require further investigations. First, the concept of complexity in the task of information gathering is important for understanding how users deal with the task and how they manage the process of gathering information on the web. The task of information gathering is more sophisticated than keyword search activities. It involves several subtasks and has a high-level goal. There are several factors that contribute to the task complexity which include: the cognitive intensity, the need for multiple sources of information, the need for different kinds of data, and the complications associated with current tools, applications, and features available for the task. Hence, understanding these factors could provide guidelines for improving tools and applications intended for the task of information gathering on the web.

Second, search during information gathering involves search activities that may include submitting search queries to web search engines, typing in and starting at a particular URL, and following links on web pages to locate information of interest. Searching is an important aspect of information gathering. In Alhenshiri, et al. (2012c), the number of search activities during the task outweighed the rest of the activities 'combined' by a noticeable difference. As Table 2 shows, all of the research studies involved showed that the task of information gathering was highly search-reliant. Therefore, it was selected as a characteristic that participates in defining the task of information gathering.

Third, information gathering is a complicated task that has a high-level goal, a characteristic that distinguishes information gathering from other kinds of tasks such as browsing information. The task goal can be either directed (e.g. learning about a topic) or undirected (e.g. searching for advice or obtaining information about a service) according to Rose and Levinson (2004). Considering the goal of the task and the user activities and behaviour based on the goal of the task may help with building tools that accommodate the needs of users performing information gathering on the web.

Fourth, treating information gathering as a combination of subtasks instead of simple activities, such as submitting a query and copying information, provides insights into understanding the strategies followed by the user and the tools used to achieve each of the subtasks. Hence, difficulties associated with the subtask can be revealed.

Improvement can be investigated for each subtask which would eventually contribute to the overall improvement of the information gathering task.

Fifth, since information gathering on the web is a complex task that involves multiple subtasks, it may require that the user works on the task over multiple sessions as shown in some of the research highlighted in Table 2. The importance of this feature comes also from the need for exploring multiple resources, different types of data and topics, and the cognitive intensity of the task that necessitates dealing with different subtasks and issues related to reasoning and decision making. Understanding the need for multiple sessions for information gathering tasks on the web could provide insights into the design of tools and features that would improve how users keep and re-find information over multiple sessions and the kind of information that should be considered for these two subtasks.

Sixth, information gathering tasks usually require the need for multiple sources of information. The choice of considering this particular characteristic was motivated by the studies shown in Table 2. The fact that information gathering is more complicated than keyword search comes from the need for exploring and filtering through multiple sources of information for reasoning and decision making. Relating information from multiple sources is needed for this type of task because information sources on the web suffer from issues of reliability and trustworthiness.

Seventh, the need for multiple data types during the task may also necessitate exploring, filtering, and searching through several websites and pages to accomplish the task goal. Understanding this characteristic of the task could lead to improving the design of tools intended for information gathering on the web. Such tools may take into account the processes of filtering, comparing, selecting, formatting, and editing the different kinds of data during the task.

Finally, having to deal with multiple sources of information, searching for and filtering through different kinds of data, and using different strategies and features to accomplish different subtasks involved in the task of information gathering on the web results in the need for multiple applications. Even though users use different applications and tools available for information gathering, these tools suffer from several drawbacks. Some of these tools are used only because of their availability and not their effectiveness or efficiency. Current available applications and tools do not suffice either for accomplishing the task requirements or for assisting users with the different subtasks and their associated strategies. Hence, choosing this characteristic could help in the design of the tools that would improve how users gather information on the web.

3.3. The Process of Information Gathering on the Web

With respect to the task process, information gathering can be viewed as a set or sequence of subtasks. A subtask is a set of activities users perform to achieve a sub-goal related to the overall goal of the task. Some of the subtasks take place in sequence, i.e. finding information sources and then finding information, while others happen simultaneously, i.e. managing and keeping information, during the gathering process. The process of completing an information gathering task involves several subtasks three of which are mandatory for each task. Information gathering must involve (at least) finding information sources, finding information, and managing information. For example, if the task is only concerned with locating a piece of information, it becomes a ‘finding’ or a ‘browsing’ task since no actual gathering takes place. The subtasks are presented in the following list.

1. Finding information sources: this subtask involves activities such as using web search engines to find web pages and sites in addition to searching by typing in a URL and searching by navigation. This subtasks is affected by the kind of search the user uses.
2. Finding information: This subtask involves searching for information on sites and pages by using the site search features or by following links on a page. Information can be of different types including text, images, tables, videos, etc. This subtask is affected by: the type of search, the kind of resources to be found, and the data types.
3. Keeping information: this subtask involves preserving and saving different forms of information using different strategies during the task of information gathering. Keeping information can be temporary (within the same session) for filtering, comparing, and selecting information for the task. Keeping information may also

happen over multiple sessions. Keeping information can be for reusing or simply re-finding. The first type involves keeping information gathered as parts of the task requirements while the second type involves keeping sources of information for further finding, comparing, or decision making. Keeping is affected by the task complexity, its cognitive intensity, the features and applications used, and the number of sessions needed for the task.

4. Re-finding information: Re-finding information usually happens in subsequent sessions to information 'sometimes' preserved earlier or information located yet was not kept (through researching). Users use different strategies for re-finding information such as searching bookmarks, emails, files, and re-searching the web (if no information was preserved). This subtask is affected by the overall complexity of the task of information gathering; the cognitive intensity of the task; the types of data needed to be gathered; the features, tools, and applications used; and the information resources that need to be explored.
5. Organizing and managing information: This subtask includes activities related to organizing the information as required in the task. Activities such as copying and pasting; formatting; and editing text, images, and other kinds of data are also considered components of this subtask. This subtask is affected by the task complexity, the search factor, the goal of the task, the types of data needed, and the applications used for the task.
6. Comparing information, reasoning, and decision making: This subtask is related to cognitive activities for selecting what is relevant for the task. It is affected by the task complexity, the required data, and the applications used.
7. Interpreting and reviewing the task: This subtask depends on the user level of background knowledge and the complexity of the task. The outcome of the task in terms of gathered information depends on those two factors.

3.4. SUMMARY

Identifying the subtasks that comprise the task of information gathering on the web and clarifying the features associated with each subtasks helps with further investigations. Each subtask can be investigated and issues users have with the subtask can be identified.

The contribution of every subtask to the overall task of information gathering can also be studied. The model developed in this chapter helped with the investigations of difficulties and issues users have with information gathering on the web by investigating the individual subtasks. Two preliminary studies that were conducted to explore the subtasks of finding information sources and finding information are discusses in the following chapter.

CHAPTER 4 PRELIMINARY STUDIES

This chapter discusses two preliminary studies that were conducted as part of this research to explore the concept of information gathering on the web. The studies were meant to examine the use of aspects of visualization and clustering in finding information and information sources, the first two subtasks of the information gathering task. The results of these two studies contributed to shaping the rest of the research journey.

4.1. VSE, A VISUAL SEARCH ENGINE FOR IMPROVING HOW USERS GATHER INFORMATION ON THE WEB

The first attempt to investigate the concept of information gathering on the web took the subtasks of finding information and information sources into consideration. The VSE was designed to investigate the effectiveness of specific visualization and clustering features on how users accomplish the task of information gathering.

4.1.1.Design and Implementation

Relevant results for user queries typically reside somewhere among the list of hits provided by the search engine. The large number of matching documents and the textual list format make it hard for users to find such results. To further enhance the effectiveness of finding information and information sources, the VSE was guided by the following motivations:

- Visualization should make best use of the display to render results of large hit sets in an easily understood form.
- The use of visualization should incorporate features of web documents such as the page size, last update, thumbnail, and the page similarity to other pages in the result set in order to provide detailed insights into the returned documents (Kules, et al., 2006).
- The design of visualization techniques should consider the kind of user and the type of task. Most web search users are casual users without specific search training (Laycock, 2009).

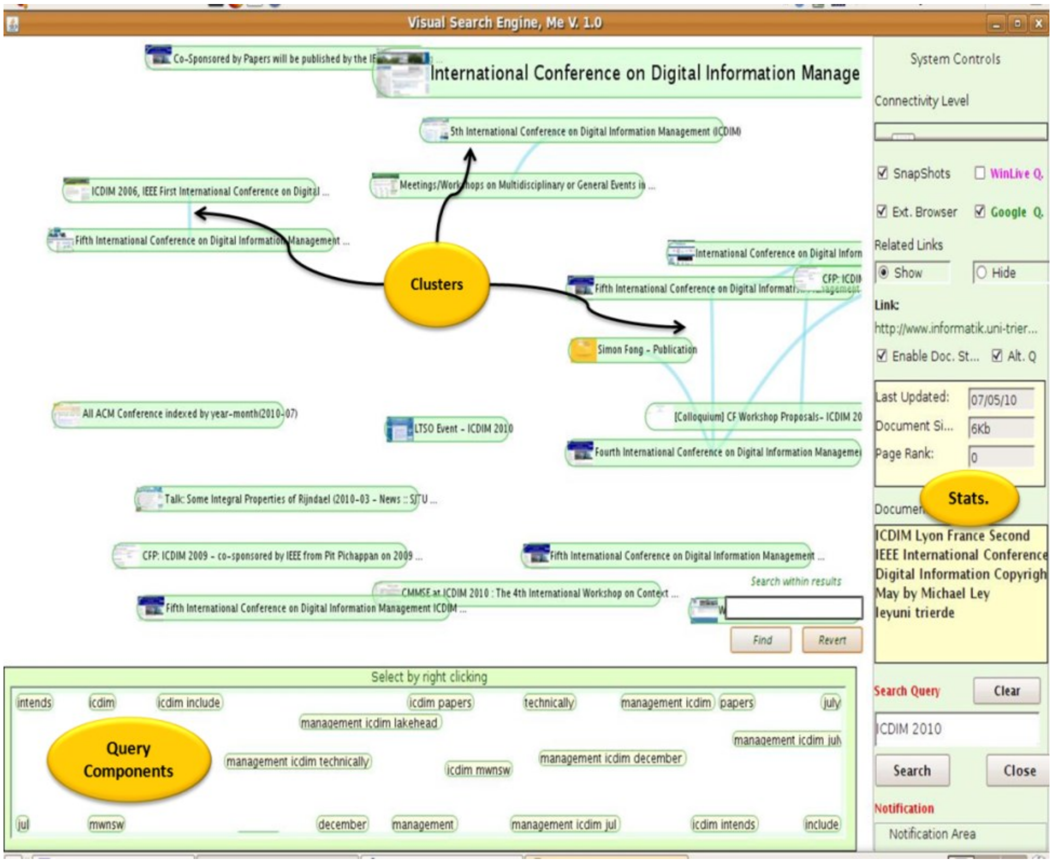
The need for better presentations of search results has become increasingly important (Jacso, 2007). To improve access to web search results and the effectiveness of users in gathering information on the web, a prototype search interface (the VSE) was designed and implemented. The design of the VSE relied on three main principles. First, the VSE used two underlying search engines, namely *Google* and Microsoft's search engine, *WindowsLive*. Second, an alternate query was generated based on the user query. A single term query was augmented with additional terms to create an alternate query using the semantic network *WordNet* (Miller, 1990; Moldovan and Mehalcea, 2000). In the case of multiple term queries, an alternate query was generated by randomly reordering the keywords of the original query. The third principle was the incorporation of a query formulation area. The VSE infers keywords and complete phrases from the search results. Those query components are presented along with the search results. The user can select terms and phrases in the reformulation area to build further queries.

4.1.2.Search and Response

Each time the user submits a new user query, an alternate query is created as discussed above. Both queries are submitted to Google and WindowsLive. The resulting documents are examined to eliminate repeated matches. Then, the derived results are further manipulated to build clusters of similar documents using the cosine similarity measure with a threshold applied to the document-document similarity matrix (Manning, et al., 2008). Document summaries are analyzed to infer keywords and phrases to be shown to the user in the query formulation area. For each document, the VSE derives a thumbnail from the publically available Google preview directory (no longer available). The VSE uses thumbnails attached to titles as recommended in the work of Teevan, et al. (2009). The system allows the user to enable or disable the use of alternate queries and page thumbnails and to set the document similarity threshold.

4.1.3.The Interactive Search Interface

The interactive interface was designed to enhance the user's ability to find relevant results (information sources) for a task. This is done based on two main principles: visualization and interactive query reformulation. On the left upper side of the VSE interface shown in Figure 10, search results are presented as visual glyphs containing document titles and web page thumbnails. Similar documents are connected via edges on the display. The VSE implements a simple content-similarity-based clustering that utilizes document summaries provided by the underlying search engine. Users can click on any glyph to either open its corresponding page in the web browser or hide the glyph from the display to reduce clutter. Moreover, the system provides drill-down capabilities. Users can hover over a glyph to magnify its content and reveal the document statistics, summary, and URL. Document statistics include size, PageRank value, and last update. The right side of the VSE interface in Figure 10 provides a view of the document statistics. Moreover, the interface offers the user the ability to search within the results.


Visual Search Engine, Me V. 1.0
Co-Sponsored by Papers will be published by the IE...
International Conference on Digital Information Manage
5th International Conference on Digital Information Management (ICDIM)
Meetings/Workshops on Multidisciplinary or General Events in ...
ICDIM 2006, IEEE First International Conference on Digital ...
Fifth International Conference on Digital Information Management ...
Clusters
International Conference on Digital Inform
CFP: ICDIM
Fifth International Conference on Digital Informati... ...agement
Simon Fong - Publication
All ACM Conference indexed by year-month(2010-07)
LTSO Event - ICDIM 2010
[Colloquium] CF Workshop Proposals- ICDIM 20
Fourth International Conference on Digital Information Manageme
Talk: Some Integral Properties of Rijndael (2010-03 - News :: SJTU ...
CFP: ICDIM 2009 - co-sponsored by IEEE from Pit Pichappan on 2009 ...
Fifth International Conference on Digital Information Management ...
CMMSE at ICDIM 2010 : The 4th International Workshop on Context ...
Fifth International Conference on Digital Information Management ICDIM ...
Search within results
Find
Revert
Select by right clicking
intends
icdim
icdim include
icdim papers
technically
management icdim
papers
july
management icdim lakehead
management icdim juh
Query Components
management icdim technically
icdim mwnsw
management icdim december
jul
mwnsw
december
management
management icdim jul
icdim intends
include
System Controls
Connectivity Level
Related Links
Stats.
Documen
Search Query
Clear
Search
Close
Notification
Notification Area

☑ SnapShots ☐ WinLive Q.

☑ Ext. Browser ☑ Google Q.

● Show ○ Hide

Link: http://www.informatik.uni-trier...

☑ Enable Doc. St... ☑ Alt. Q

Last Updated: 07/05/10

Document Si... 6Kb

Page Rank: 0

ICDIM Lyon France Second IEEE International Conference Digital Information Copyrigh May by Michael Ley leyuni trierde

ICDIM 2010

Figure 10. The VSE interface.

The second principle is interactive query reformulation. Along with search results, the VSE presents users with alternate query terms and phrases as shown on the lower part of the VSE interface in Figure 10. The VSE selects terms and phrases from the document summaries returned by the underlying search engines. The terms and phrases are presented to the user in a separate view on the display. The interface provides immediate search and response. It allows the user to enable and disable several parts such as document statistics, document thumbnail, edges representing content similarities, and the display of results derived for either the alternate query or the original one. In addition, users can see the actual content of a page in the web browser. As a result, users with different query building skills can easily use the VSE.

4.1.4.Research Questions

The study conducted to evaluate the visualization and clustering features embedded in the VSE was intended to find answers to the following questions:

1- With regard to finding information sources and information for the given task, how effective is the VSE compared to the conventional search engine Google in:

    a. Increasing the number of relevant pages found?

    b. Increasing the number of topics (types of information) covered in the pages located?

    c. Minimizing the frequency of query submission to find information sources on the web?

    d. Minimizing the need for opening web pages to find information on information sources?

2- From the user's perspective, how enjoyable is the VSE compared to Google while searching for finding information sources and finding information on the web?

4.1.5.VSE Evaluation

To evaluate the VSE, a user study was conducted. The study investigated the ability of the VSE to improve the effectiveness in performing information gathering tasks on the web. Kellar, et al. (2007) classified web tasks as fact-finding, information gathering, and navigation. In both fact finding and navigation tasks, the documents sought by the user are usually specific and they require some previous knowledge about the document content. Therefore, the results are usually either relevant or non-relevant. Information-gathering tasks, on the other hand, achieve results that depend on the user's search expertise. This type of tasks may involve searching for information that belongs to different topics. This study investigated four main factors: the ability of users to find relevant results effectively and efficiently, the number of topics (types of information) covered in the results, the number of queries needed by the user, and the number of pages the user opened. In the study, the VSE was compared to Google, which is considered to be the most frequently used search engine according to Search Engine Watch[10].

4.1.6.Study Location and Population

The user study was carried out in the Faculty of Computer Science at Dalhousie University using a Linux machine loaded with the Chromium web browser and the VSE. The Chromium web browser was chosen because of its ability to record the data required in the study. Fourteen participants took part in this study. The participants were computer science students including five undergraduate and nine graduate students. Nine of the participants were males and five were females.

4.1.7.Study Design

The design of the study was complete factorial in which all combinations of the tasks and the two systems (Google and VSE) were used. The study design was within-subjects and counterbalanced so that each participant experienced all conditions in different random orders. This design limits the effect of order.

4.1.8.Study Tasks

The search tasks were information gathering intended to encourage participants to find as many pages related to the task as needed. They did not reflect all the information gathering tasks in reality. The following is an example of the tasks used in the study. It contains several unnecessary restrictions, and it is very limited due to time issues since all the participants were volunteers.

"Use the given search system to gather web pages that include information about how to use the java programming *language in transforming html documents into images. The pages you find should give someone a good idea about the task's topic. You can submit up to five queries only, and you should not go beyond one page of results for each query you submit. You can also view results in the web browser."*

4.1.9.Study Methodology

To ensure fairness in the comparison between the VSE and Google, WindowsLive was not used during the study and the number of results was restricted to 20 hits per query for

[10] http://searchenginewatch.com/

both the VSE and Google. Each participant received a short training session using one demonstration task on the VSE. Then, each participant was asked to perform their information gathering task using both the VSE and Google. The two search engines had the same chance of being used first to eliminate any learning effect. After performing the search task, each participant was asked to report the number of relevant results and the number of topics (or types of information) covered by those pages. Each participant was asked to state their confidence level (on a Likert scale) with the results.

After performing a task on the VSE and another task on Google, each participant was asked to complete a post-study questionnaire. In this questionnaire, the participants were asked to answer questions in which the VSE and Google were compared. Finally, participants were asked to provide their own comments regarding the VSE and any possible improvements from the participant's point of view. The participants spent a total of 20 minutes on average to complete the study.

4.1.10. Study Data

The data collected in the study included logged data as well as questionnaires' data. The log data included the time on task and the number of pages opened on the web browser using both the VSE and Google. While participants were working on their tasks, the machine logged all search queries. The questionnaire data included answers to questions in the questionnaires. The questions were used to compare the VSE to Google with respect to the ease of use, the effectiveness with respect to finding relevant pages, and the effectiveness of the query reconstruction feature as well as other visual features of the VSE as perceived by the participants. The questionnaire data was collected using a five-point Likert scale (1 for worst and 5 for best). The analysis of the data considered answers of 1 and 2 as negative choices and answers of 4 and 5 as positive. Finally, the data involved qualitative comments reported by the participants.

4.1.11. Study Results

4.1.11.1. Efficiency

With respect to efficiency, the task completion time was less using the VSE with an average task time of 6.5 minutes compared to Google with an average of 8.2 minutes.

Although the difference is not statistically significant, there is a good indication that the more familiar the users become with the VSE, the faster they may perform.

4.1.11.2. Effectiveness (Quantitative Performance Results)

Effectiveness was measured with regard to the number of relevant pages located for the task on each search engine, the number of topics (information types) discovered in the results, the number of pages participants had to open on the browser (click behaviour), and the number of queries they had to submit in order to achieve the task. The VSE was shown to be more effective than Google with respect to those criteria. The quantitative results are shown in Table 3. The ANOVA test results showed a significant difference between the VSE and Google ($F(1,12)=41, p <0.003$) with respect to all of the above evaluation metrics. Figures 11, 12 and 13 provide illustrations of the differences between the two search tools.

Table 3. Quantitative results.

System		***VSE***	***Google***
Time (mean***)***		*6.5*	*8.2*
Submitted queries	Mean	2.5	3.5
	SD	1.5	1.6
Pages opened on the browser	Mean	2.1	9
	SD	1.5	11
Relevant pages found	Mean	6.5	4.5
	SD	2.7	2.1
Covered topics	Mean	3	2
	SD	1.4	1.3

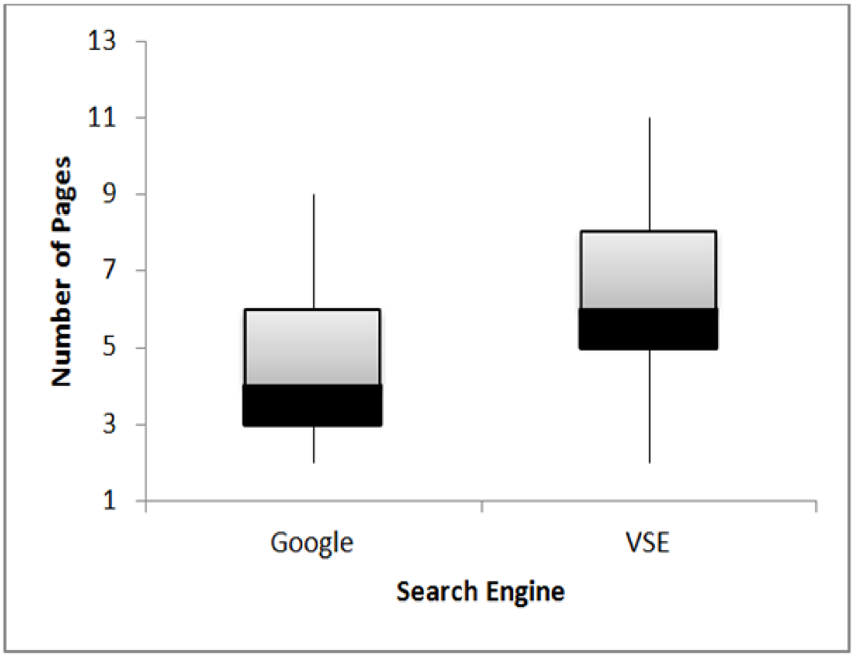


Figure 11. Pages located for the task.

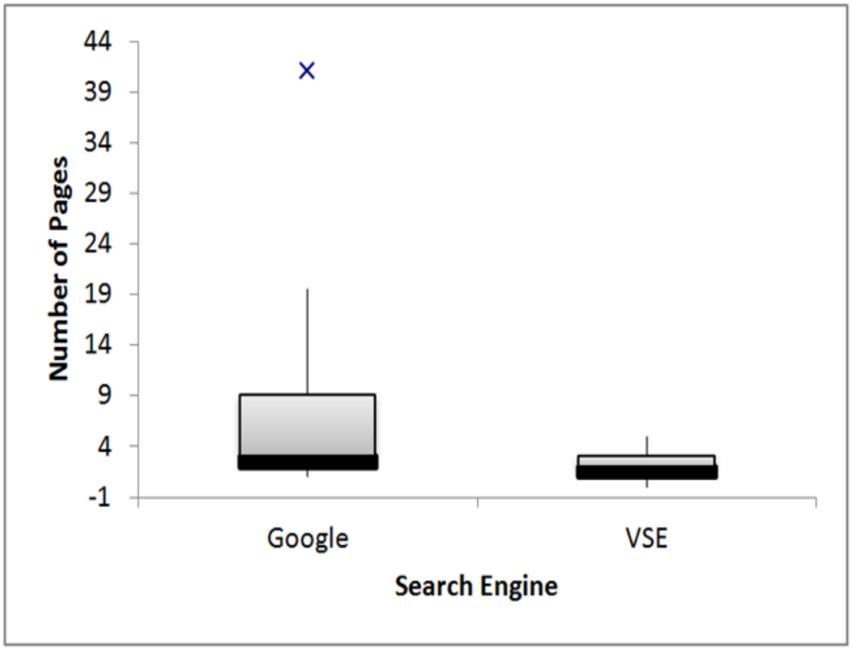


Figure 12. Pages opened on the browser during the task.

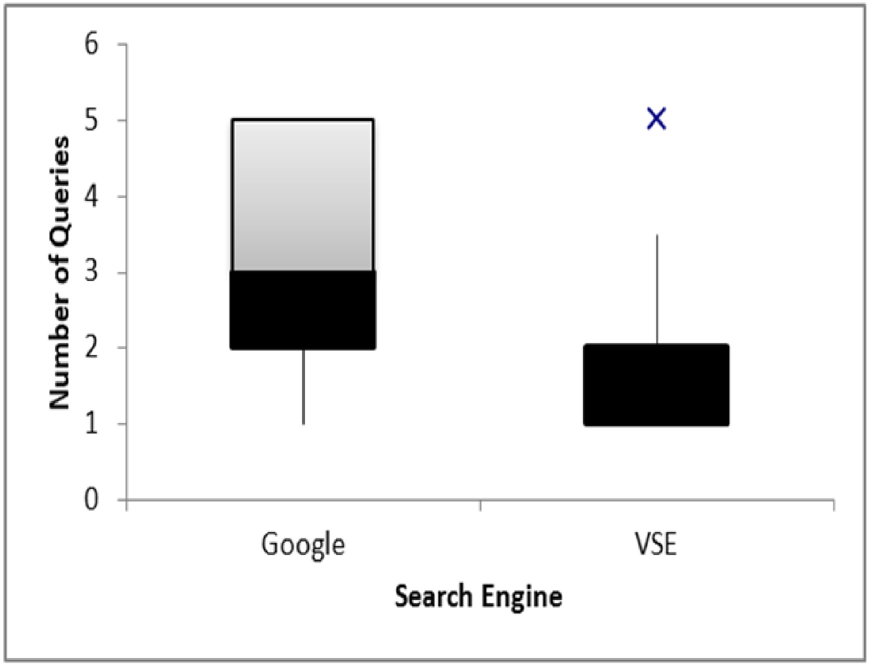


Figure 13. Queries submitted for accomplishing the task.

4.1.11.3. Effectiveness (Questionnaire Results)

Regarding the user confidence with the results, the study showed that the participants were more confident with their choices of pages using the VSE than they were with Google. Figure 14 shows the effectiveness of the VSE compared to Google with respect to user confidence. In addition, analyzing the data of the questionnaires showed that 65% (9/14) of the participants reported that the VSE was better than Google, that the VSE made it easier to find the intended results, and that it made query reformulation more efficient. Finally, 71% (10/14) of the participants considered the VSE as a more helpful search tool. Figure 15 shows the subjective ratings. A *z*-test showed no significant difference between the two proportions of ratings of the VSE and Google (*z =1.42, p = 0.08*).

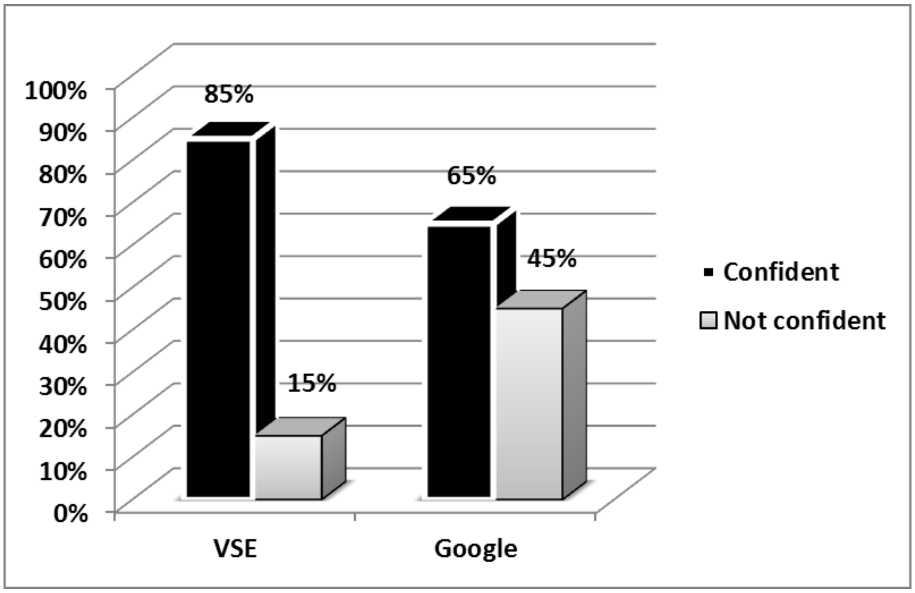


Figure 14. A comparison of the user confidence with the VSE and Google.

4.1.11.4. Effectiveness (Subjective Comments)

The last part of the results was the qualitative comments provided by the participants. Participants provided 40 comments, which were categorized by the researchers into four types namely: comments related to the query reconstruction area, comments about the visual view of the results, comments regarding the visual clustering used, and comments related to finding relevant results. The comments regarding the VSE were 80% positive. The first set of comments concerned the query reconstruction aspect of the VSE. All of those comments were positive. The visual view of the results showed that 71% (10/14) of the participants regarded the VSE as better for viewing web search results. The negative comments expressed concerns with the result presentation layout and the slow movement of the glyphs on the display. One participant stated that "*the VSE keeps me focused and interested in what I am doing, but I do not like the slow movement of the connected glyphs*". According to the *z*-test, there was a significance difference between the two proportions of comments ($z = 2.79, p < 0.004$).

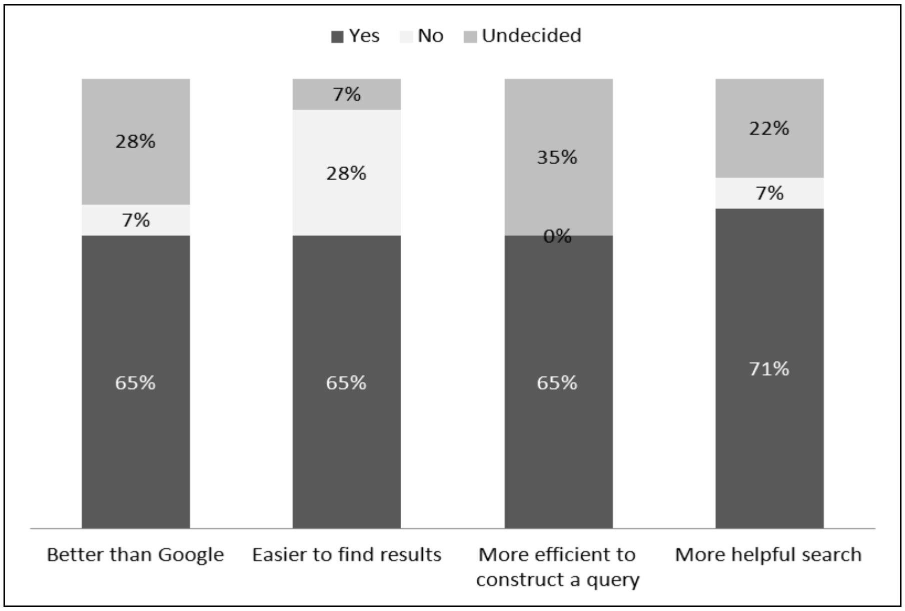


Figure 15. Post-study questionnaire ratings of the VSE compared to Google.

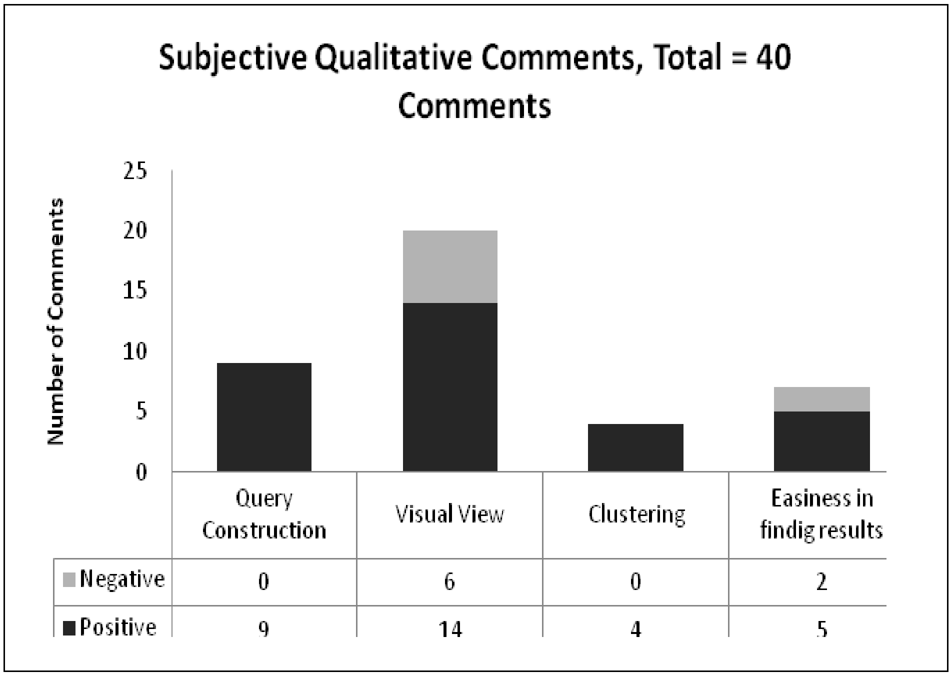


Figure 16. Subjective comments.

Clustering documents based on their content similarity received all-positive comments. The last group of comments was on the easiness in finding relevant results. Only 28% of the qualitative comments were negative One of the participants stated, "*I would like to set and learn more about everything in the VSE so that I can use it in a more effective way*".

The qualitative comments indicated that the VSE was more efficient with respect to allowing its users to quickly spot relevant documents on the display without having to open each document in the web browser. Figure 16 shows the distribution of the qualitative comments among the four groups.

4.1.12. Study Limitations

Due to limited resources, the study was conducted with the participation of computer science students only. In addition, the number of participants was limited to only fourteen individuals who may reflect early adopters. Also, the choice of information-gathering tasks does not permit for generalizing the study results to all information gathering tasks.

4.1.13. Discussion

The study, while small, showed that presenting web search results visually accompanied with query reconstruction components (keywords and phrases) and document information required submitting fewer queries, viewing fewer pages on the browser, and spending less time on the task. In addition, the exploitation of document features provided by the VSE's underlying conventional search engine including the page URL, title, summary, thumbnail, and document statistics in addition to the visual clustering of similar pages lead to improving the effectiveness in gathering information on the web.

The effectiveness of the VSE was based on the number of pages found by the user for the task, the number of topics (or types of information) in those pages, the number of queries submitted, and the number of pages opened by the user in the web browser. Utilizing visualization in addition to providing different features of the search results along with query construction components created a better environment for the user to find relevant results for the task.

Furthermore, that the view of the search results found in the VSE was more effective than the presentation of Google was demonstrated in the questionnaire results and the qualitative user comments. For example, the participants' ratings favored the VSE presentation over Google's. In addition, participants were more confident in their decisions regarding the task results with the VSE than they were with Google. The qualitative comments showed a significant difference between the two tools with regard to the query construction and results presentation approaches.

Previous research has investigated visualization and clustering in web results presentation and query formulation. However, most of the evaluations have been concerned with the effectiveness of the search technique in simple query-response contexts. In web information gathering, the user may have to gather different sources of information, find further related information to the located sources, and re-find and reuse information that was previously located (Broder, 2002; Rose and Levinson, 2004, Kellar, et al., 2007). The implementation of the VSE took the subtasks of finding information sources and finding information into consideration. The visual view included document statistics, categorization of similar documents, and query components. Consequently, users found it more effective to find information and information sources using the VSE than using Google.

4.1.14. Conclusion

An interactive search interface (the VSE) that utilized multiple underlying search tools to provide visual search results, web document statistics, and query reconstruction components was presented. A user study was conducted to evaluate the effectiveness of the VSE. The results of the study showed that the VSE model for presenting web search results improved the effectiveness of users gathering information on the web.

4.2. VLN, A VISUAL LINK NAVIGATION INTERFACE FOR FINDING INFORMATION WITHIN WEB HIERARCHIES

After investigating the subtasks of finding information sources and finding information in the previous study, the research went a step further to tackle the issue of finding information within websites and pages as a continuation of the subtask of finding information. The VLN was meant to assist users gathering information from the web to locate information within web hierarchies by navigation.

4.2.1.Design and Implementation

A prototype interface, VLN, was built using a tree layout for navigating through websites given a starting point (a URL). The interface, shown in Figure 17, was designed to provide the following features:

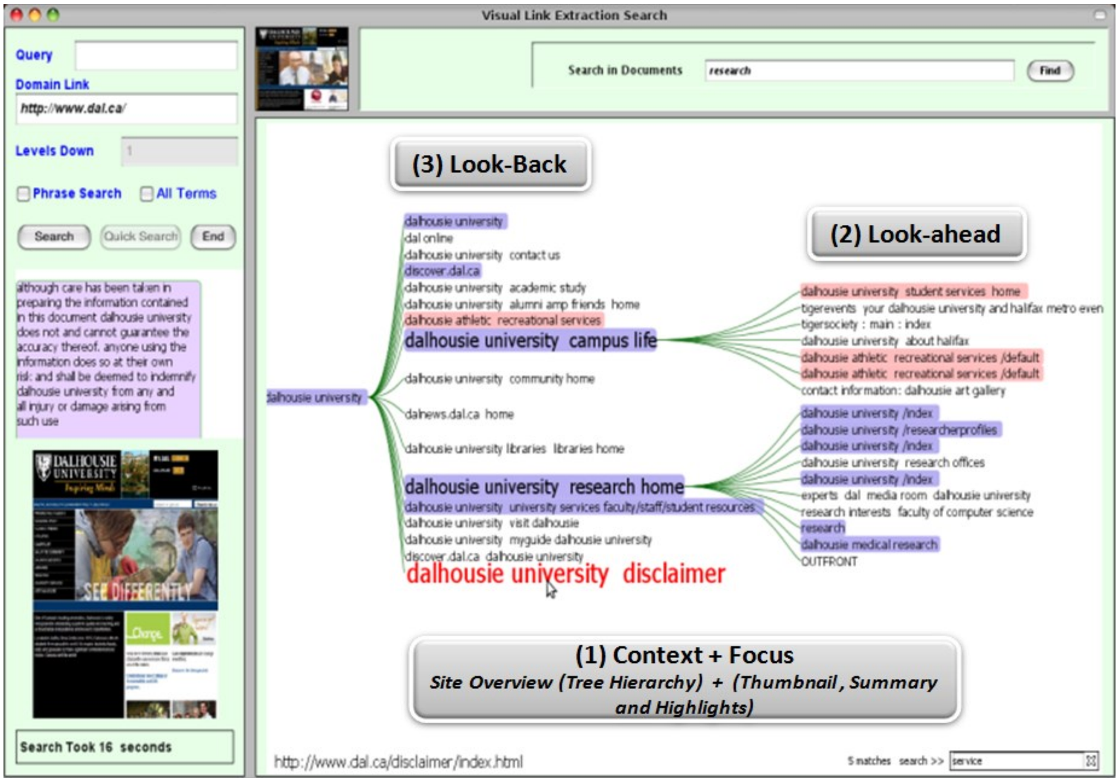
a. Context and Focus. The presentation of the hierarchy of hyperlinks provides context to users navigating while searching for information. Focus is provided by the on-demand extension of the tree nodes that represent web pages.

b. Look-ahead. The on-demand expansion of the tree nodes and the hover-over feature that allows the user to see the summary and thumbnail of the page minimize the time spent reading text during navigation.

c. Look-back. The visual presentation of the page connections, with previously visited pages highlighted, provides a simple means for the user to return to those pages for comparisons.

It is the intent of the VLN to improve the experience and effectiveness of users gathering information by minimizing the time spent reading pages during navigation, reducing the need for query formulation and reformulation within a website, and allowing users to focus on the task at hand.

The VLN works by extracting hyperlinks embedded in the page that represents the starting point of navigation. For each extracted link, the VLN downloads the page content and the page thumbnail. Then, it creates the page summary by eliminating HTML tags and extracting the human-readable text on the page. The page title and URL are also extracted. The VLN parses the major types of files found on the web such as HTML documents, text documents, PDF documents, and so forth. Processing tree levels continues to the limit chosen by the user. The VLN continues to display the tree hierarchy while processing of further levels happens behind the scene. Each time the user clicks on a tree node, the same extraction process applies.



Visual Link Extraction Search

Query

Domain Link http://www.dal.ca/

Levels Down 1

□ Phrase Search □ All Terms

Search in Documents research

although care has been taken in preparing the information contained in this document dalhousie university does not and cannot guarantee the accuracy thereof. anyone using the information does so at their own risk: and shall be deemed to indemnify dalhousie university from any and all injury or damage arising from such use

Search Took 16 seconds

http://www.dal.ca/disclaimer/index.html

5 matches search >> service

Search
Quick Search
End
Find
(3) Look-Back
(2) Look-ahead
(1) Context + Focus
Site Overview (Tree Hierarchy) + (Thumbnail, Summary and Highlights)
dalhousie university
dalhousie university
dal online
dalhousie university contact us
discover.dal.ca
dalhousie university academic study
dalhousie university alumni amp friends home
dalhousie athletic recreational services
dalhousie university campus life
dalhousie university community home
dalnews.dal.ca home
dalhousie university libraries libraries home
dalhousie university research home
dalhousie university university services faculty/staff/student resources
dalhousie university visit dalhousie
dalhousie university myguide dalhousie university
discover.dal.ca dalhousie university
dalhousie university disclaimer
dalhousie university student services home
tigerevents your dalhousie university and halifax metro even
tigersociety : main : index
dalhousie university about halifax
dalhousie athletic recreational services /default
dalhousie athletic recreational services /default
contact information: dalhousie art gallery
dalhousie university /index
dalhousie university /researcherprofiles
dalhousie university /index
dalhousie university research offices
dalhousie university /index
experts dal media room dalhousie university
research interests faculty of computer science
research
dalhousie medical research
OUTFRONT
DALHOUSIE UNIVERSITY
Inspiring Minds
SEE DIFFERENTLY

Figure 17. The VLN interface, root URL (www.dal.ca).

The VLN was created using Java swing components and the *prefuse* visualization toolkit discussed in the work of Grewal, et al. (2000). The interface provides three modes of access to web pages from a given website. That is a given web page is used as the root of a virtual website defined by out-links. The first mode is the tree view of the website for users to navigate within the site by clicking or expanding nodes in the tree. The user can control the number of levels used in constructing the tree. By default, the link visualization process stops at the next level. Hyperlink repetition within the tree is eliminated.

Second, thumbnails of pages, shown in Figure 17, that the user already visited within the site, derived from Google preview, are available for direct revisiting. The third mode is by queries, including phrase matches, generated by the user for searching within the current website. The interface has two search features. The first one is for searching through the titles of the web pages in the tree. Crowded trees benefit the most from this search feature. The second feature allows for searching in the text of the documents. Figure 17 shows both kinds of search results highlighted in different colors.

By using the VLN interface, the user can potentially see hundreds of web pages connected to the root of the tree representing the website. Previously viewed pages are bolded as shown in Figure 17. The user can zoom in and out to control the viewable context on the tree. In addition, to facilitate navigation in this space, the VLN provides three features. First, the user can search the tree to locate pages that contain keywords in the body or title of page. Second, a text chunk of each page is provided on the left sidebar of the interface to help users look ahead at information in pages. Finally, thumbnails are provided to stimulate the user's mental images of the web pages.

4.2.2.Research Questions

The research study conducted to evaluate features embedded in the VLN prototype was intended to find answers to the following questions:

1. Compared to the ordinary web browser, how effective is the VLN in:
    a. Finding relevant information to the task in web page hierarchies through search by navigation?
    b. Minimizing the depth required for navigating through website hierarchies?
    c. Locating more consistent results for the task?
2. Compared to the ordinary web browser, how enjoyable is the VLN to users searching by navigation to find information on the web?

4.2.3.VLN Evaluation

Of particular interest in this study was the effectiveness of this approach for finding information in web pages as part of information gathering , a type of task that represents around half of the user tasks on the web. Our concern, in this case, is increasing the effectiveness of the user in identifying appropriate information in the pages linked to a given website. That is, the second step of most information gathering tasks after finding the information source (site or page). To evaluate the VLN for information gathering within websites, a small user study was conducted. The study examined effectiveness, success in finding pages with information required in the task, efficiency in completing the task, and user preference compared to a regular browser. The results of the study were promising and showed that the visualized tree-based approach was more effective, more efficient, and preferred by users in this study.

4.2.4.Study Location and Population

The evaluation study took place in the Web Information Filtering Laboratory at the Faculty of Computer Science, Dalhousie University. Ten participants, all students, from the Faculty of Computer Science at Dalhousie University took part in this study. Seven of the participants were males while three were females. The study was conducted using a computer running the Linux platform and equipped with the Chromium web browser along with the VLN. The Chromium web browser was selected because of its ability to record data required in the study.

4.2.5.Study Design

The design of the study was complete factorial in which all combinations of the values of task and process (the independent variables in the study) were involved. The study was within-subjects and counterbalanced so that participants experienced each of the conditions but in different orders. This design limits the effect of order and the effect of sequence of tasks on processes.

4.2.6.Study Tasks

The authors designed two information-gathering tasks within a given site for the study. While the tasks are similar, both the design of the two websites and the familiarity of the users with the websites differed. The students had at least a working knowledge of the university website and little, if any, experience with the bank website. The bank website was much flatter and less complicated than the university site. Each task asked the participants to find four pages containing information related to the request stated in the task.

Task 1: "Suppose that you have a friend who asked you to find out about the kind of services TD Canada Trust Bank provides in order to open an account and do other businesses with this bank such as having a credit card, a loan, or buying an investment. Perform navigation starting from the TD main web page (http://www.tdcanadatrust.com) to achieve your task. You are required to find four pages which you can send to your friend, and which you feel they would give clear ideas to your friend about their request. Write down the URLs for the selected pages you find most appropriate for your friend's request."

Task 2: “Suppose that you have a friend who asked you to find out about the study programs including tuition fee, living expenses, recreational facilities and all possibly useful pages at Dalhousie University. Perform navigation starting from the Dalhousie main web page (http://www.dal.ca) to achieve your task. You are required to find four pages which you can send to your friend and which you feel would give clear ideas to your friend about their request. Write down the URLs for the selected pages which you find most appropriate for your friend's request.”

4.2.7.Study Methodology

Each participant in the study was given a short training session on the VLN. Following the training session, each participant completed both tasks using both the VLN and the Chromium browser. The order of tasks and the choice of interface were completely counterbalanced. At the end of each task, the participant was asked to complete a short questionnaire that recorded whether the participant found what they were looking for and how confident they were in their choices of web pages for the task. After finishing all tasks, each participant completed a post-study questionnaire. Participants spent a total of 25 minutes on average to complete the study.

4.2.8.Study Data

During the study, the following data were collected: time on task, pages visited during the task, maximum depth of the site hierarchy reached by the user on the VLN and the browser, and data collected in the per-task and the post-study questionnaires.

4.2.9.Study Results

4.2.9.1.Learning Effect

A learning effect may have resulted from asking the participants to repeat the tasks on both systems. We found that independent of which system or task participants used first, 80% of the users were confident in their results, the average navigation depth was 1.7 levels on both interfaces, there was little difference in the time required to finish the tasks, using the browser took 23 minutes on average, and using the VLN took 24 minutes. An ANOVA test confirmed that there was no significant learning effect ($f(1,8)=0.15, p<0.71$).

4.2.9.2. Task Effect

An ANOVA test was performed between the tasks on both the VLN and the browser. The results (*f(1,8)=1.2, p=0.26*) indicate that the choice of task did not have a significant effect on the participants performance.

4.2.9.3. Navigation Path

We were interested in knowing how many levels participants explored down the website hierarchy while searching for information to complete their task. The participants explored down fewer levels with the VLN than they did with the browser to accomplish the tasks. Figures 18 and 19 show the normalized navigation depth for the university and the bank tasks. The results indicate that participants needed to drill down about half as far on the university website using the VLN (Mean=0.7 levels, SD=0.39) as they did using the browser (Mean=1.7 levels, SD=0.19). Similarly, for the bank task, participants explored deeper (Mean=1.8 levels, SD=0.18) using the browser than they did using the VLN (Mean=1.3 levels, SD=1.05). The difference between the VLN and the browser in both tasks was significant using ANOVA (*F(1,8) =45, p<0.0002*). We anticipate that the smaller difference in the bank case between the VLN and browser can be related to the shallower structure of the bank website.

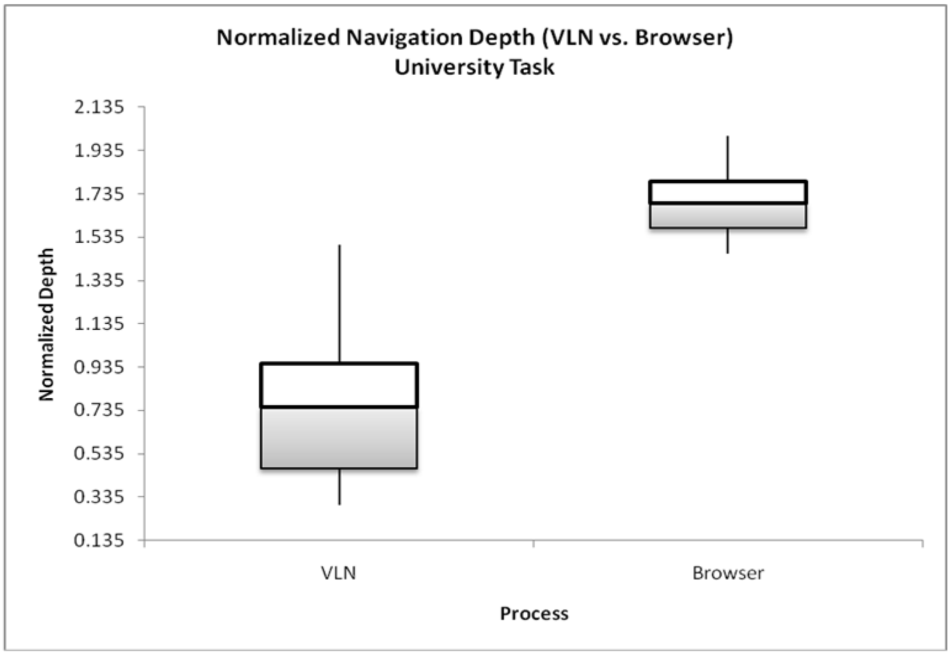


Figure 18. Task 1 (University) navigation results.

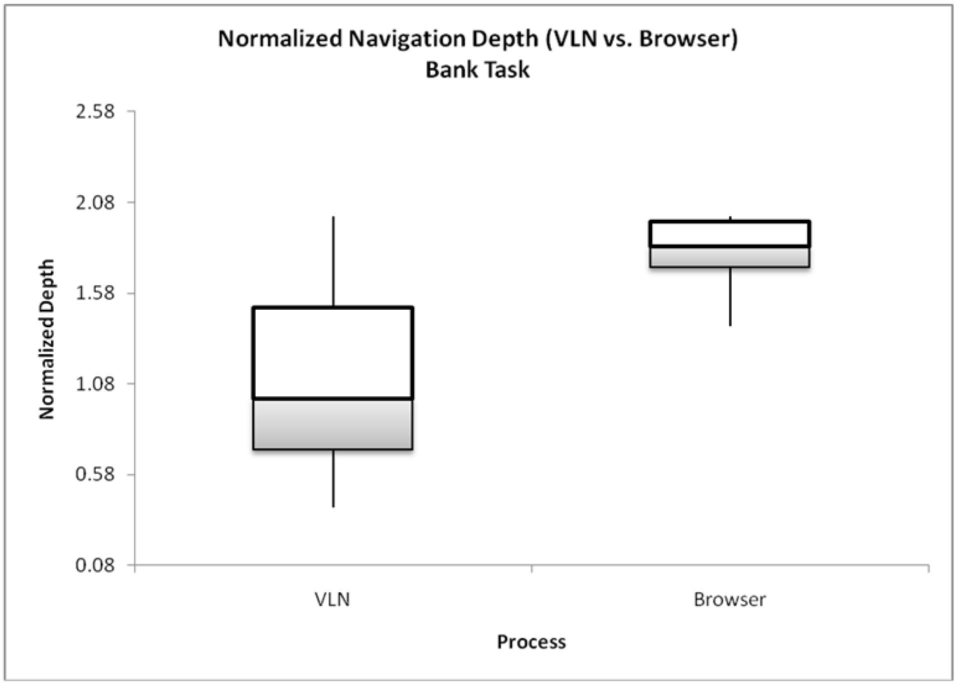


Figure 19. Task 2 (Bank) navigation results.

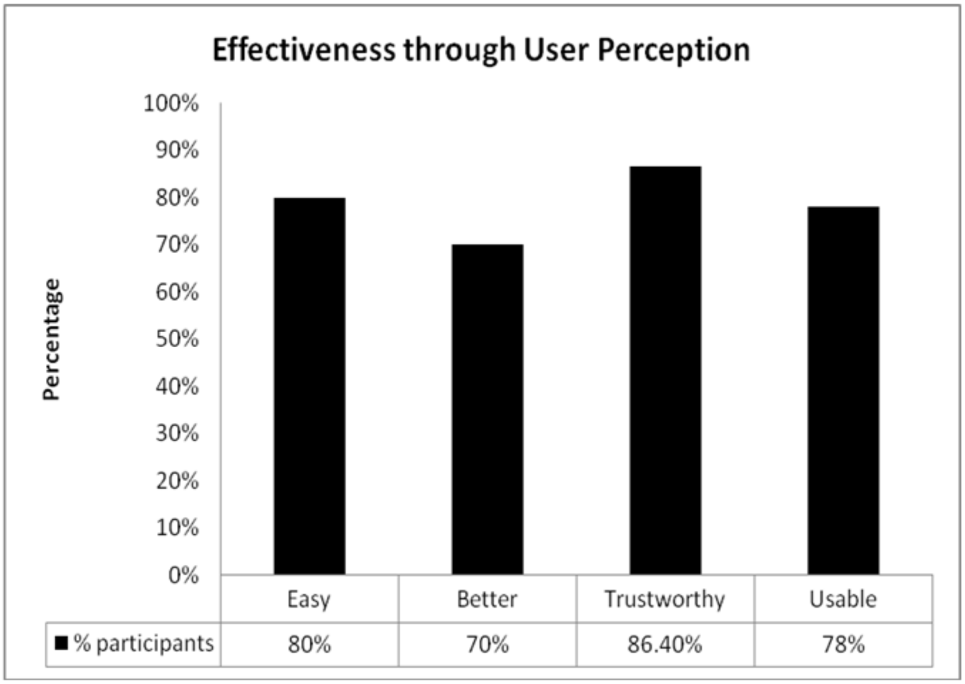
4.2.9.4. Task Results

To investigate the effectiveness of the VLN with regard to accomplishing the tasks, we analyzed several factors. First, we examined the number of unique pages selected by the participants for each task using each system. Overall, participants were more consistent in their page selections for the tasks when they used the VLN. Over both tasks using the VLN, a total of 40 pages were selected by the users. Of those, 10 were unique, indicating 4 selections per page. However, in the case of the browser, a total of 37 pages were selected. Of those pages, 14 were unique, indicating 2.6 selections per page. When we consider only the bank task, 18 pages in the 40 selected (2.2 selections per page) were only chosen once. The ANOVA test results show a significant difference between the VLN and the browser with respect to page selections (*F(1,8) =4.7, p< 0.034*).

Second, we examined the page overlap between the VLN and the browser. The overlap in pages chosen by using the two systems was low; six pages for the university task and nine pages for the bank task. We surmise that the low overlap in pages chosen by the users between the VLN and the browser reflects the personal differences and the openness of the tasks. We also anticipate that the higher overlap in the case of the bank task may be related to the flatter structure of the site itself.

4.2.9.5. User Perception

To investigate the user perceptions of the VLN effectiveness, the study considered several measures including the ease of use of the interface, the intention for future use by the participants, the participants' confidence in the results, and the participants' preference of the VLN over the browser. Figure 20 shows the results of these measures. For 80% of the participants, the VLN was easy to use. It took them three minutes on average to be able to use it. Moreover, 70% of the participants thought that the VLN interface was better than the browser in expressing the hierarchy of the website, and it helped them to make choices of pages to view next. Of the participants, 86% indicated that they were more confident using the VLN in navigating the websites to locate task results. Finally, 78% of the participants stated that they would use the VLN in the future. The data was collected using a five-point Likert scale (1 for worst and 5 for best). The analysis of the data considered answers of 1 and 2 as negative choices while answers of 4 and 5 as positive. The *z*-test results showed a significant difference between the proportion of users who considered the VLN as effective and those who considered the browser as effective (*$z = 2.05$, $p < 0.03$*).



	Easy	Better	Trustworthy	Usable
■ % participants	80%	70%	86.40%	78%

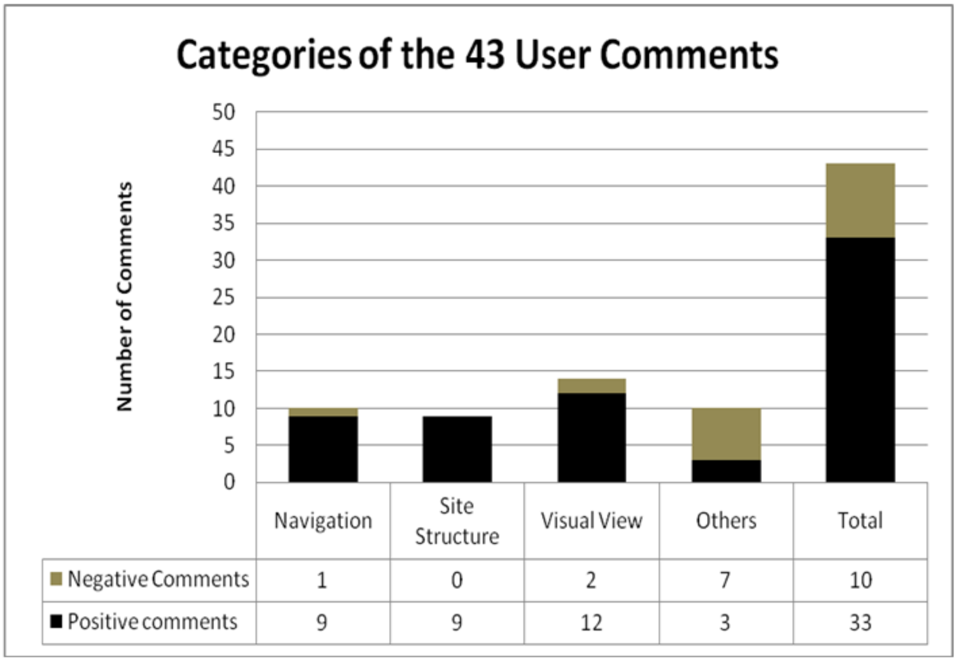
Figure 20. Subjective perceptions of the VLN.

4.2.9.6. Qualitative Comments

The post-study questionnaire provided qualitative data related to the effectiveness of the VLN. The 43 user comments were grouped under four categories as shown in Figure 21.

The first set of comments regarded the use of the VLN for navigation. The users rated the VLN as an excellent navigation tool in 90% of the comments. The second set of comments regarded the use of the tree structure for representing websites on the VLN interface. All participants stated that the site structure was clear, and useful. An example of a user comment is "*The view of the site on the VLN allowed me to make fast decisions about the search direction so I did not have to waste time looking at the page content.*" Most participants stated that they preferred being able to see the entire view of the site through the hierarchical presentation of hyperlinks. A significant difference between the two proportions of comments was revealed using the ***z***-test (*z= 3.13, p < 0.002*).

Regarding remembering the navigation path, 85% of the participants considered the use of the tree structure as helpful or very helpful. Only two comments (15%) considered following navigation paths on a crowded tree in the VLN layout to be hard. The last set of comments concerned other interface features. For example, users requested bigger views of the thumbnails and reduced clutter in the case of dense websites. One user stated, "*Although the view was interesting and kept me focused on the job, I found it hard to make use of the presented thumbnail images*".



	Navigation	Site Structure	Visual View	Others	Total
▨ Negative Comments	1	0	2	7	10
■ Positive comments	9	9	12	3	33

Figure 21. Recorded user comments.

4.2.10. Discussion

The results of the study indicate that the features emphasized in the prototype: context, look-ahead, and look-back, were important to the user while navigating for information gathering. The tree visualization of hyperlink connectivity provided a context for participants to gain the big picture of the part of the web graph in focus (represented by the virtual website). This is clear when considering their options and information found on pages. The tree visualization and the highlighted features (thumbnail and summary) by hovering over the tree nodes allowed the users to look ahead before committing to opening a page and to return to pages that are more important. Participants explored fewer levels of the site hierarchy using the VLN than they did using the browser. Moreover, among the users in the study, the VLN was more consistent with respect to the results located for each task. We interpret the lower number of selections per page in the case of the web browser as due to the hidden hierarchy of the site in the current browsing model.

The qualitative results provide an indication that users might prefer this approach for some tasks of information gathering. Participants indicated that the VLN was excellent for navigating websites while searching for information. It was easy to use and comprehensive in the way it expressed the structure and content of websites. Furthermore, participants stated that they were very confident in their results using the VLN interface because they were able to see their navigation paths at all times, and they were able to see the content, title, hyperlink, and thumbnail of the page in advance.

Some participants raised concerns about the size of the thumbnails and issues of clutter on the tree view. Although clutter can be avoided at least partly by the use of queries within a website complemented by hierarchical navigation of links, this issue may be investigated in future work. Overall, the user study, while small, indicates that showing the hierarchy of a website is useful for information gathering tasks once a website is known. Furthermore, we speculate that, a model for information gathering that combines query formulation and navigation through embedding hyperlinks in the body of the web page may provide significant benefits.

4.2.11. Study Limitations

This was a preliminary study with a small population of only ten participants. We recognize that having all of our participants from the Faculty of Computer Science does not reflect the population of Internet users but does reflect early adopters. In addition, the restriction to two websites and two tasks is a limitation of the applicability of our results to the broader web. Finally, since it was difficult to hide the identity of the VLN, the user decision of what system to favor might have been slightly biased. However, considering the performance results shown above limits the effect of such bias.

4.2.12. Conclusion

In this study, a technique for information gathering within websites was presented. The study was meant to investigate the effects of the VLN on gathering information from websites (the subtask of finding information) as part of the information gathering task on the web. The results of the study indicate that the VLN interface provided both positive user perceptions and improved effectiveness benefits to users navigating websites while searching for information. Future research may involve investigating variant views of the hierarchy of websites, varied search tasks, and larger and more diverse populations.

4.3. SUMMARY

The studies presented in this chapter showed that the use of aspects of visualization and clustering improved the effectiveness of how users accomplished the task of information gathering. In particular, improving the subtasks of finding information and information sources was considered in the design of the tools examined. The studies' results showed that other subtasks such as re-finding information for the task of information gathering and managing and organizing the information gathered for the task require further investigations. The users in the studies showed concerns with re-finding the task information, keeping track of web pages, and organizing the information gathered. The subtasks of re-finding information and managing and organizing information are further considered in the studies discussed in Chapters 5 and 6.

The studies discussed in this chapter resulted in altering the model initially created for the task of information gathering. The subtasks of finding information and finding related information were combined since the core aspect of these two subtasks is locating information that satisfies the task requirements. All the information that needs to be located for the task is essentially related to the task topic. In addition, the user activities for keeping information are considered to contribute to the subtask of managing and organizing information. Re-finding information is considered to have connections to the three core subtasks of finding information sources, finding information, and managing and organizing information. Re-finding can happen for already located sources of information or parts of the information. It can also happen to parts of the task that were kept during the managing process.

All in all, the information gathering task model was modified by: adding the identified features of the task to each applicable subtask, changing the model to reflect the core subtasks, and better reflecting the relationships between subtasks in the information gathering task on the web. The modified model is shown in Figure 22.

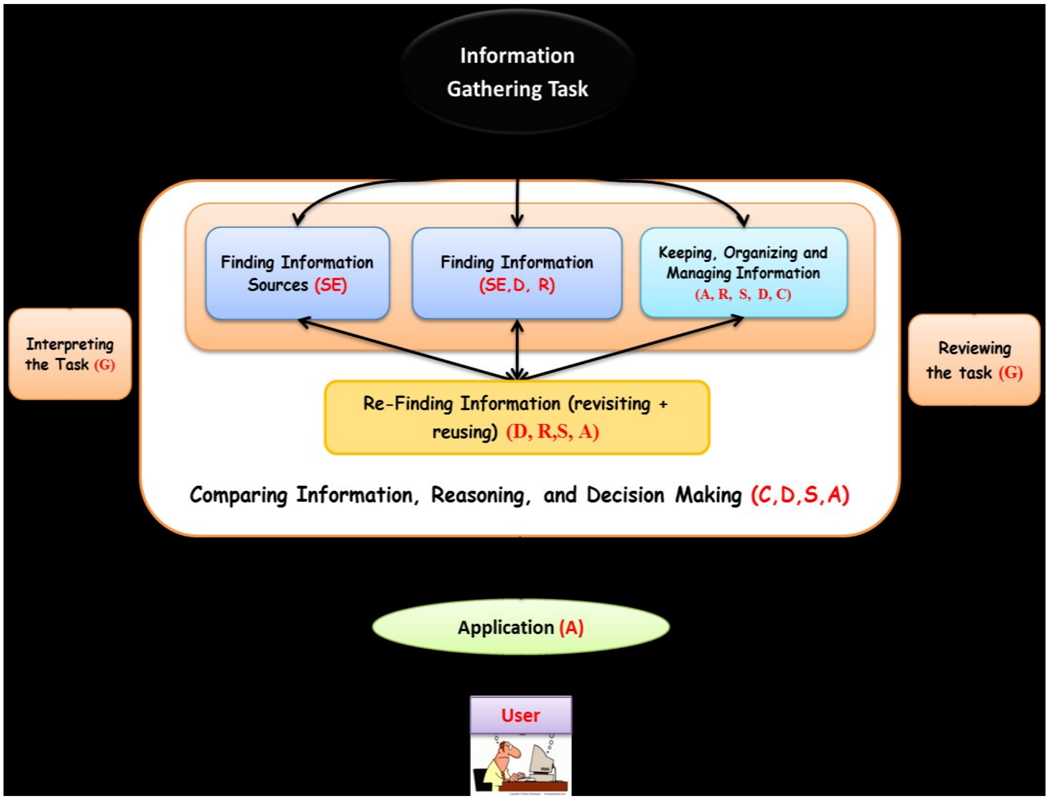

Information Gathering Task
Finding Information Sources (SE)
Finding Information (SE,D, R)
Keeping, Organizing and Managing Information (A, R, S, D, C)
Interpreting the Task (G)
Reviewing the task (G)
Re-Finding Information (revisiting + reusing) (D, R,S, A)
Comparing Information, Reasoning, and Decision Making (C,D,S,A)
Application (A)
User

Figure 22. The modified model of the information gathering task.

CHAPTER 5 INVESTIGATING THE TASK OF INFORMATION GATHERING ON THE WEB

5.1. RESEARCH STUDY

The study this chapter discusses was conducted to investigate how users manage and organize information and how they re-find information during the task of information gathering on the web (see Figure 22). The study examined difficulties users encounter and tools and strategies they use while managing, organizing, and re-fining information for the task. Possible improvements to the process of information gathering on the web were also considered in the investigation.

5.1.1.Research Questions

To provide design recommendations that would help the implementation of information gathering tools and features for the web, the study intended to answer the following research questions:

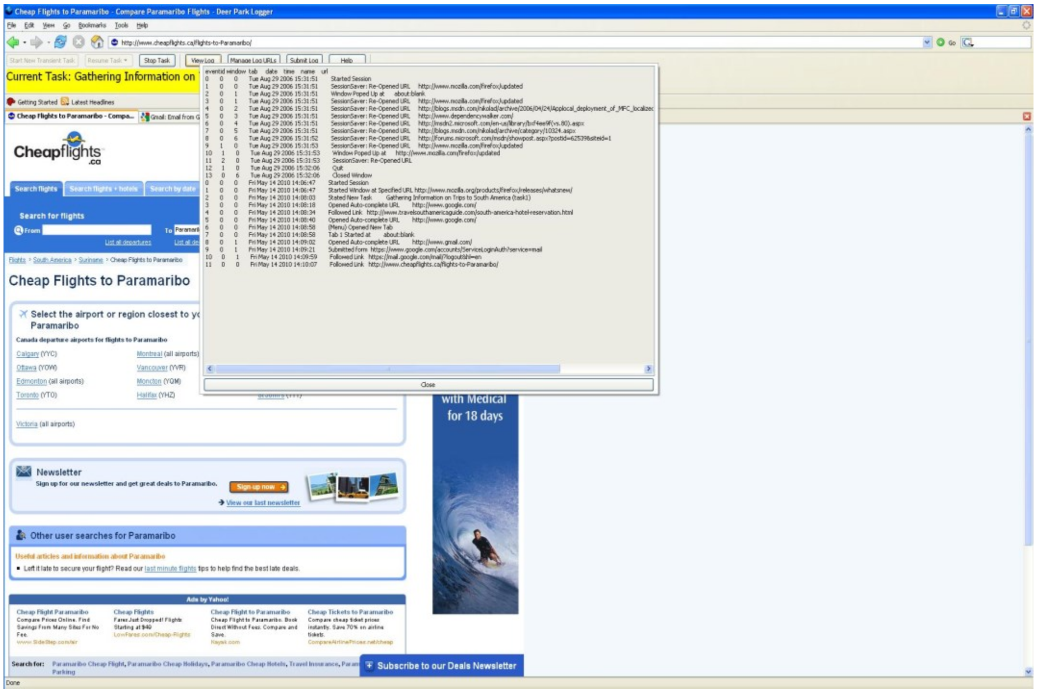
1. What activities do users perform during information gathering tasks and at what capacity?
2. What tools and strategies do users use to manage, organize, and re-find the task information?
3. What difficulties do they encounter while managing, organizing, and re-finding the task information?
4. What recommendations could be drawn for future design of tools intended to be used for gathering information from the web?

5.1.2.Study Design and Population

The design of the study was complete factorial and counterbalanced. There were 20 participants in the study of which 10 were graduate students and the remaining 10 were undergraduate students. All participants were students from the Faculty of Computer Science at Dalhousie University. The study used a special version of the Mozilla Firefox browser[11] called *DeerParkLogger* (Figure 23), which was built at Dalhousie University. This browser has the ability to log all activities users perform during the task such as

[11] http://www.mozilla.com

submitting search queries, opening URLs in the browser, copying and pasting information, and the like. Every participant in the study was assigned two different tasks. Each task had the same chance of being used first. The study tasks are discussed next.


Cheap Flights to Paramaribo - Compare Paramaribo Flights - Deer Park Logger
Start New Transient Task
Resume Task
Stop Task
View Log
Manage Log URLs
Submit Log
Help
Current Task: Gathering Information on
Cheapflights.ca
Cheap Flights to Paramaribo
Close
Subscribe to our Deals Newsletter

```
eventid window tab   date   time   name   url
0   0   0   Tue Aug 29 2006 15:31:51   Started Session
1   0   0   Tue Aug 29 2006 15:31:51   SessionSaver: Re-Opened URL   http://www.mozilla.com/firefox/updated
2   0   1   Tue Aug 29 2006 15:31:51   Window Poped Up at   about:blank
3   0   1   Tue Aug 29 2006 15:31:51   SessionSaver: Re-Opened URL   http://www.mozilla.com/firefox/updated
4   0   2   Tue Aug 29 2006 15:31:51   SessionSaver: Re-Opened URL   http://blogs.msdn.com/nikolad/archive/2006/04/24/Applocal_deployment_of_MFC_localizec
5   0   3   Tue Aug 29 2006 15:31:51   SessionSaver: Re-Opened URL   http://www.dependencywalker.com/
6   0   4   Tue Aug 29 2006 15:31:51   SessionSaver: Re-Opened URL   http://msdn2.microsoft.com/en-us/library/bxf4ee9f(vs.80).aspx
7   0   5   Tue Aug 29 2006 15:31:51   SessionSaver: Re-Opened URL   http://blogs.msdn.com/nikolad/archive/category/10324.aspx
8   0   6   Tue Aug 29 2006 15:31:52   SessionSaver: Re-Opened URL   http://forums.microsoft.com/msdn/showpost.aspx?postid=62539&siteid=1
9   1   0   Tue Aug 29 2006 15:31:53   SessionSaver: Re-Opened URL   http://www.mozilla.com/firefox/updated
10  1   0   Tue Aug 29 2006 15:31:53   Window Poped Up at   http://www.mozilla.com/firefox/updated
11  2   0   Tue Aug 29 2006 15:31:53   SessionSaver: Re-Opened URL
12  1   0   Tue Aug 29 2006 15:32:06   Quit
13  0   6   Tue Aug 29 2006 15:32:06   Closed Window
0   0   0   Fri May 14 2010 14:06:47   Started Session
1   0   0   Fri May 14 2010 14:06:47   Started Window at Specified URL http://www.mozilla.org/products/firefox/releases/whatsnew/
2   0   0   Fri May 14 2010 14:08:03   Stated New Task   Gathering Information on Trips to South America (task1)
3   0   0   Fri May 14 2010 14:08:18   Opened Auto-complete URL   http://www.google.com/
4   0   0   Fri May 14 2010 14:08:34   Followed Link   http://www.travelsouthamericaguide.com/south-america-hotel-reservation.html
5   0   0   Fri May 14 2010 14:08:40   Opened Auto-complete URL   http://www.google.com/
6   0   0   Fri May 14 2010 14:08:58   (Menu) Opened New Tab
7   0   0   Fri May 14 2010 14:08:58   Tab 1 Started at   about:blank
8   0   1   Fri May 14 2010 14:09:02   Opened Auto-complete URL   http://www.gmail.com/
9   0   1   Fri May 14 2010 14:09:21   Submitted form   https://www.google.com/accounts/ServiceLoginAuth?service=mail
10  0   1   Fri May 14 2010 14:09:59   Followed Link   https://mail.google.com/mail/?logout&hl=en
11  0   0   Fri May 14 2010 14:10:07   Followed Link   http://www.cheapflights.ca/flights-to-Paramaribo/
```

Figure 23. *DeerParkLogger*, the version of Mozilla Firefox that was used in the study.

5.1.3.Task Description and Construction

The study used four different information gathering tasks each of which had two parts (e.g. Task 1_a and Task 1_b). The reason for splitting each task into a sequence of two related parts was to provide a context in which participants might find some advantage in re-finding information for Task1_b that was found or kept during Task1_a. The tasks developed for the purpose of the study were simulations of realistic tasks users would usually perform for gathering information on the web. Each task was created following the guidelines described in (Kules and Capra, 2008). Those guidelines are summarized in the following:

- The task description should indicate: uncertainty, ambiguity in information need, or need for discovery.
- The task should suggest knowledge acquisition, comparison, or discovery.
- It should provide a low level of specificity about the information required in the task and how to find such information.

- It should provide enough imaginative contexts for the study participants to be able to relate and apply the situation.

A focus group of sixteen graduate students was used to ensure that the tasks were of the same level of complexity. The focus group was given the tasks to discuss and to decide on the degree of complexity for each task based on the following criteria:

- The time needed to complete the task.
- The amount of required information to be gathered in the task.
- The clarity of the task description.
- The possible difficulties that the user may encounter during the task.

The focus group had two meetings. In the first meeting, the group discussed the difficulties associated with the tasks. During the first meeting, the discussion revealed that two of the four tasks were more difficult than the other two tasks. The group suggested changes in the requirements of the two more-difficult tasks. The principal investigator applied those changes to the requirements and descriptions of the tasks. In the second meeting, the focus group suggested minor changes to the tasks. Those changes were made during the meeting and the whole group agreed that the four tasks were quite similar with respect to the four criteria indicated above. The tasks used in the study are described in the following section.

5.1.4.Information Gathering Tasks Used in the Study

The following are the tasks used in the study. These tasks require finding and organizing information during the gathering process. They also state clearly the need for finding information, keeping information, re-finding information, comparing information and sources of information (web pages), and deciding on information. They suggest acquisition and discovery of knowledge. For example, Task 4 implies that the user should acquire knowledge about the status of farms in two provinces in Canada and compare these two provinces in terms of the decrease in the number of farms. Moreover, all tasks provide low level of specificity indicating the need to find and decide on the information gathered for the task. The tasks used in the study provide enough contexts for the users to apply the situation described in the task since those tasks are simulations of realistic ones.

TASK 1

a. You have a friend who asked you to provide her with valuable information about Canadian universities that she may consider for a graduate degree in business. What kind of information would you like to send to your friend providing her with a comparison of two universities? Provide your choices of the universities. Provide links to at most five web pages you find helpful in making your choices. Also, provide a copy of the information you would send to your friend, which shows the comparison you made. You will need to come back to reuse the information you found in this task.

b. Last time, you selected two Canadian universities for your friend to pursue a graduate degree in business. Now, your friend asked you to provide her with two choices of American universities to consider for a graduate degree in business. Choose two American universities that provide graduate degrees in business. Find up to five web pages that would allow you to make a comparison between the Canadian universities you already selected and the two American universities you will choose. Provide the results of your comparison (information you used in the comparison) in addition to links to at most five web pages you find useful in making the comparison.

TASK 2

a. You heard your friends complaining about bank account service charges in Canada. You are not sure why they are complaining. You want to do some research on the web to find out more about bank account service charges. State your opinion about the charges (whether your friends' complaints were true or not). Using two Canadian banks, provide links to up to five web pages that have information that you used to form your opinion. You will need to come back to the information and pages you collected in this task.

b. After you found out about bank service charges in Canada, you want to find out about bank service charges in the US. Use the information you gathered in the previous task to compare service charges of banks in Canada to those of banks in the US. Provide a comparison of service charges applied by a bank in the US to charges found in one of the two Canadian banks you identified earlier. Provide links to up to five web pages you used in comparing the two banks in addition to the information you gathered for the comparison.

TASK 3

a. While planning for a trip to South America you want to visit two countries. You want to select the two countries based on the cost of plane tickets, cost of residence, and living expenses. Provide links of up to five pages you used to make your choices. Provide the names of the two countries and a copy of the information that helped you to make the trip plan. Later, you will need to come back to the information and pages you collected in this task.

b. After planning your trip to South America, you decide that you want to make a change and visit two countries anywhere in the world. Choose two other countries and compare the countries you selected for your trip in the previous task to these two countries. State whether you should stick with the old plan or switch to the new plan based on the cost of the plane tickets, living, and residence. Please provide links to up to five web pages you used to make your decision. Also, provide a copy of any particular information that helped you with your choices.

TASK 4

a. Somebody told you that the number of farms in Nova Scotia is decreasing. On the web, you decide to look for factors behind the drop in the number of farms in Nova Scotia. How accurate is what you heard? Provide links to at most five pages that contain information you found useful. Also, provide a copy of the actual information you located on those pages. You will need to come back to the information and web pages you found for this task.

b. Now, we would like you to compare the situation in Nova Scotia to that of the province of Ontario. On the web, find out about the status of farms in Ontario. Does the situation regarding the drop of the number of farms in Nova Scotia apply in Ontario? Why/Why not? Provide links to at most five web pages you found useful for comparing Nova Scotia to Ontario. Also, provide any information you find particularly helpful in the comparison.

5.2. STUDY METHODOLOGY

Every participant was randomly assigned two tasks, each of which consisted of two parts. After being introduced to the study and signing the informed consent, every participant read the first part of both tasks. The user then performed the first part of the first task followed by the first part of the second task. Then, the user took the initial questionnaire. The time the user took to complete the questionnaire was intended to form a time gap after which the user has to re-find information kept during performing the first parts of both tasks in the second parts of those tasks. After completing the pre-study generic questionnaire, the user finished the second part of the first task. Then, the user completed the post-task questionnaire for the first task. This questionnaire created a time gap for the second part of the second task which the user performed after completing the questionnaire. The user then completed the post-task questionnaire for the second task. After the user completed the tasks and the questionnaires, the principal investigator conducted a short interview with every participant, see Figures 24 and 25.

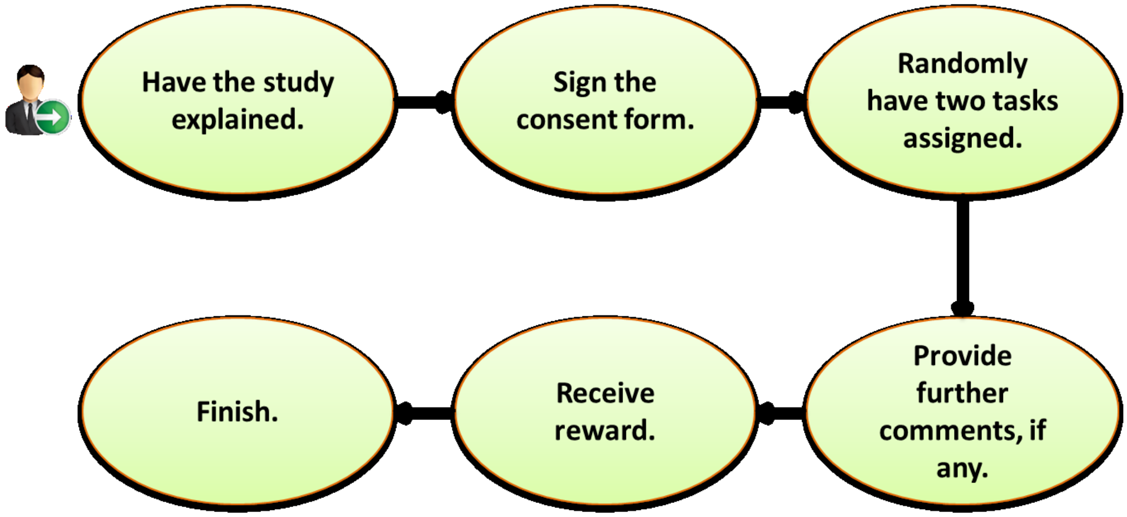

Have the study explained.
Sign the consent form.
Randomly have two tasks assigned.
Provide further comments, if any.
Receive reward.
Finish.

Figure 24. Study main steps.

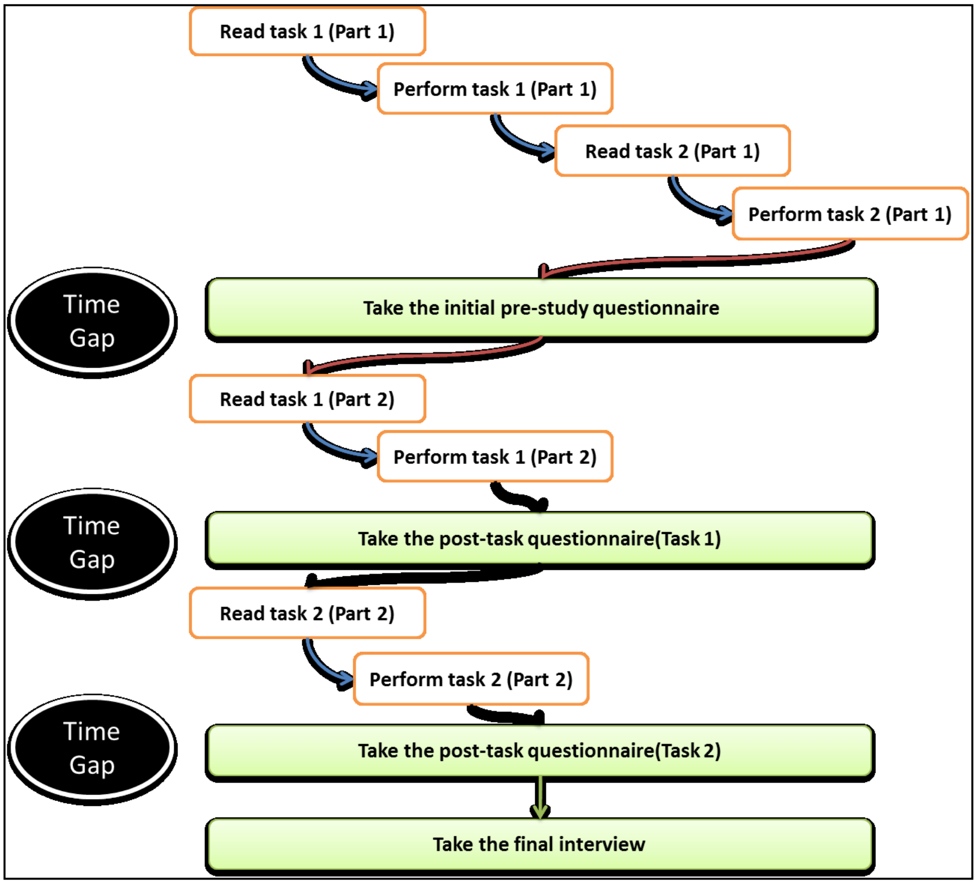


Figure 25. Study procedure.

Two questionnaires, a generic pre-study questionnaire and a post-task questionnaire were used in the study. The pre-study questionnaire asked participants about their age, their experience using the web and web search engines, their understanding and experience with web information gathering tasks, and difficulties they usually encountered while gathering and organizing information on the web. The post-task questionnaire asked participants about their understanding of the task at hand, the difficulties encountered, the level of completion they achieved with the task requirements, their confidence in completing the task requirements, and the tools the user utilized to finish the task. The interview questions covered qualitative issues such as the reasons for using certain tools and any additional comments added by the participant. Appendix A provides the pre-study questionnaire. The post-study questionnaire is shown in Appendix B. The interview questions are in Appendix C.

5.3. STUDY RESULTS

There are several factors that were taken into consideration for analyzing the process of information gathering. In particular, the study intended to investigate the subtasks of managing and organizing information and re-finding information during gathering tasks. In addition, the investigations examined how users handled multiple sessions for the given tasks. Handling multiple sessions is a subtask that emerged as part of the information gathering task as a result of this study. The investigation covered tools used during the task of information gathering, strategies users utilized to accomplish the task, and difficulties they encountered with those tools and strategies.

Information management and organization activities that the investigations considered involved the use of: browser tabs, search feature on web pages, find-on-page feature, use of bookmarks, copying and pasting information from web pages, typing information for the task requirements, and using tools in addition to the web browser. Activities involved in re-finding information and handling multiple sessions included: re-opening bookmarks, re-opening saved pages, and re-opening files and emails created during the first part of the task. Furthermore, the study considered the overall number of activities each participant performed for accomplishing the task and the time on task.

5.3.1.Demographics and Users Experience

The age of 80% (16/20) of the participants was in the range of 18-24 years. The remaining 20% (4/20) fell in the range of 25-35 years. With respect to their experience with using the web and web search engines, 80% (16/20) indicated that they were experienced users while 20% (4/20) stated that their experience was moderate. With regard to the user knowledge of web information gathering, 90% (18/20) indicated that they had good knowledge of web information gathering tasks. Only 10% (2/20) stated that they sometimes performed information gathering on the web. Users in the study have been using the web for 10.8 years on average (SD=2.9).

Of the participants, 45% (9/20) indicated that they find organizing information for gathering tasks to be difficult. Thirty percent (6/20) indicated that they were unsure while only 25% (5/20) stated that they usually find it easy to organize web information during gathering tasks. There was no significant difference between the group of users who stated that organizing the task information was easy and those who indicated that it was difficult according to the z-test results (*z=0.99, p=0.16*). In addition, 75% (15/20) indicated that they usually need multiple sessions to finish an information gathering task. The remaining 25% (5/20) indicated that they sometimes or rarely need more than one session. The difference was significant between the two proportions according to the z-test results (*z=2.84, p<0.03*).

5.3.2.Task Effect

To further ensure the equivalence of the tasks to each other with respect to the level of complexity, the differences among the study tasks were measured. The analysis used the user confidence in the task results, the time spent on the task, and the degree of difficulty in the tasks as perceived and stated by the participants. The analysis intended to investigate the possibility that with regard to any of these three factors, the tasks may have been different. The results in the three categories indicated that there was no significant difference among the tasks used in the study. The ANOVA test results demonstrated this finding (user confidence, *f(18,1)=1.97, p<0.14*; task time, *f(18,1)=0.98, p=0.33*; task perceived difficulty, *f(18,1)=0.97, p<0.42*). As concluded by the focus group prior to the study, the tasks were equivalent with respect to the time needed for the task, the requirements of the task, the complexity of the task, and the task description.

5.3.3.Time on Task

To further ensure the validity and adequacy of the tasks designed for the study based on the principles described in Section 5.1.3, the time on task was measured. Participants spent an average of 55.5 minutes on both parts of the task (SD=13.4). Table 4 shows the actual times the participants took for each task and the total study time. The tasks were designed to take approximately one hour to be accomplished. The focus group, which discussed the similarity of the tasks, attempted to make the requirements satisfiable within one hour. The difference between the time taken to complete the first task (29 minutes on average) and the time taken to complete the second task (26.5 minutes on average) was not significant according to ANOVA (*f(18,1)=0.98, p=0.33*).

Table 4. Time taken by participants for the tasks (in minutes).

Participant	Task Part (a)	Task Part (b)	Total Time
P1	24	11	35
P2	24	13	37
P3	23	14	37
P4	26	18	44
P5	25	22	47
P6	25	23	48
P7	20	28	48
P8	24	24	48
P9	31	20	51
P10	34	20	54
P11	37	20	57
P12	28	31	59
P13	20	39	59
P14	29	32	61
P15	28	35	63
P16	32	32	64
P17	37	29	66
P18	35	36	71
P19	40	34	74
P20	39	49	88
Mean	29	26.5	55.5
SD	6.2	9.7	13.4

5.3.4.User activities

To identify the different kinds of activities the users perform while managing and organizing information for the task, each instance of each activity was recorded in the log file. Participants performed an average of 125.2 activities (SD=41.9) to accomplish both tasks in the study, ranging from 42 to 198 activities per user (see Figure 26). Those activities included: opening URLs, using URL auto-completion, submitting search queries, following links on pages, bookmarking, using the find-on-page browser feature, using browser tabs, using browser windows, copying and pasting information from web pages into files and emails, typing information for the task, and using the browser history. In the second session (part) of each task, users were asked to use the information from the first session. The activities are shown in Table 5.

Table 5. User activities during the study.

Participant	Type-in URL	Submit Query	Follow Link	Bookmark	Use Browser Tab	Find on page	Copy and Past	Type Information	Use of Browser History	Auto-complete URL	Other activities	Overall activities
P1	6	1	11	5	0	0	0	1	0	0	19	43
P2	17	0	42	2	57	2	4	1	0	2	16	143
P3	5	3	53	2	5	4	16	0	0	0	31	119
P4	0	23	84	4	0	0	22	1	0	0	49	183
P5	0	8	90	5	39	2	17	1	0	18	19	199
P6	2	13	41	0	19	2	7	0	2	2	16	104
P7	1	19	46	0	35	6	20	2	0	1	7	137
P8	3	8	34	0	33	9	17	1	0	5	11	121
P9	1	8	76	0	18	0	2	0	0	0	13	118
P10	13	15	28	0	22	2	7	2	0	4	20	113
P11	8	15	77	3	47	2	6	0	0	1	11	170
P12	3	4	2	50	0	1	15	1	0	4	31	111
P13	3	10	63	0	47	0	2	2	0	0	2	129
P14	4	0	57	0	13	0	10	2	0	3	15	104
P15	3	5	22	0	38	0	11	1	0	3	13	96
P16	0	4	22	0	2	12	4	1	0	1	20	66
P17	2	8	49	0	16	0	10	0	0	0	35	120
P18	2	7	61	0	19	1	36	2	0	2	26	156
P19	1	6	60	0	85	1	19	2	0	2	25	201
P20	6	9	25	0	6	0	16	0	0	0	9	71
Total	**80**	**166**	**943**	**71**	**501**	**44**	**241**	**20**	**2**	**48**	**388**	**2504**
Mean	*4*	*8.30*	*47.15*	*3.55*	*25.05*	*2.20*	*12.05*	*1*	*0.10*	*2.40*	*19.4*	*125.2*
SD	*5.38*	*6.16*	*25.46*	*11.08*	*22.53*	*3.27*	*8.72*	*0.79*	*0.45*	*3.99*	*10.99*	*41.92*

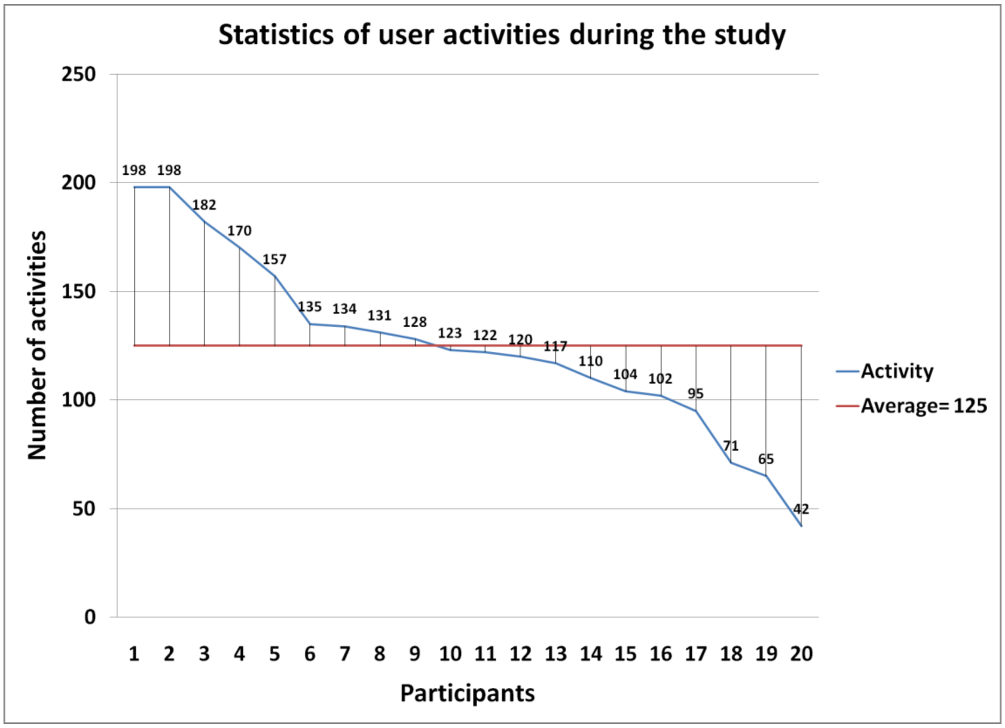


Figure 26. User activities during the study.

Table 6. Information gathering activities performed in the study.

Subtask	Activity	Frequency of activity
Managing and organizing information	Using a browser tab	501
	Using a browser window	37
	Typing information for the task	20
	Copying and pasting	241
Re-finding information	Bookmarking	71
Finding information sources + Finding Information	Submitting a query	161
	Finding information on a page	43
	Auto-completing a URL	41
	Typing in a URL	80
	Following a link	416
	Searching on a site	10

A subset of those activities (see Table 6) was selected for further examination. Figure 27 shows the frequency of each of these activities during the tasks. The results indicate that among the activities related to managing and organizing information during web information gathering tasks are using browser tabs and copying and pasting information. Those activities represented a large portion (45%) of the total number of activities indicating the high frequency of managing and organizing activities during gathering tasks.

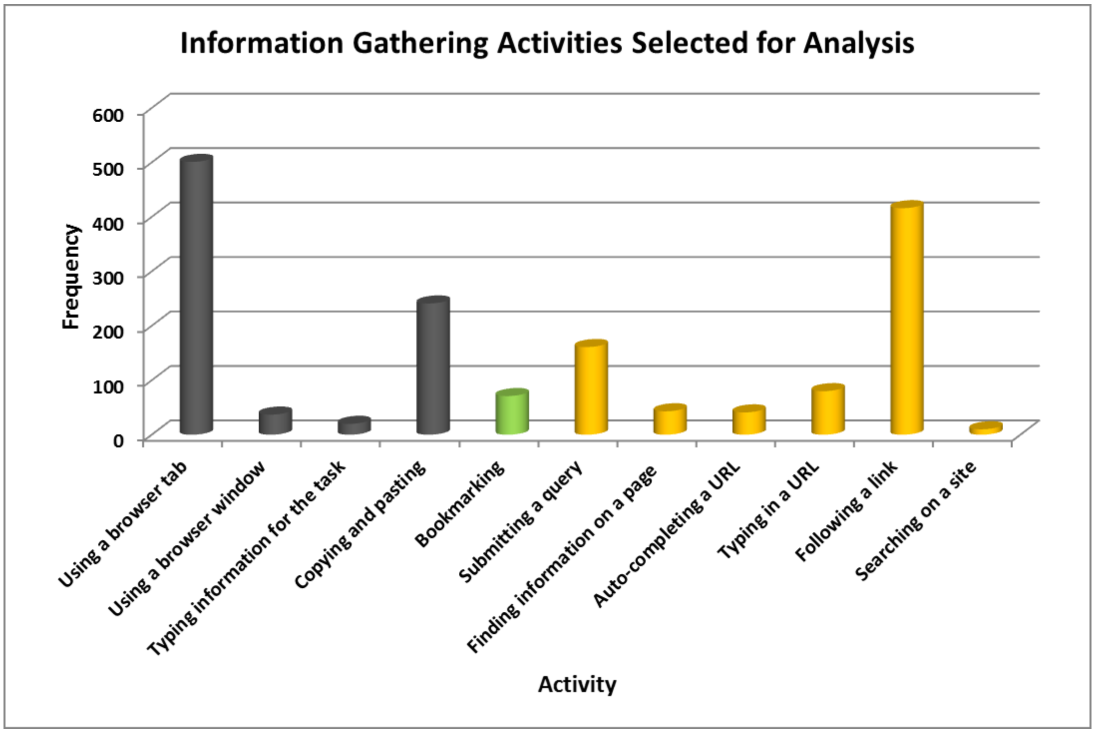


Figure 27. Information gathering activities.

5.3.5.Organization Tools

Prior to performing the tasks, users were asked to complete a pre-study questionnaire in which they stated the kind of tools they usually use to organize web information during gathering tasks. The results showed that users make use of text editors, browser features, emailing, and less frequently pen and paper. These and other tools were given individually in the participants' answers to the questions in the questionnaires and were later categorized by a group of eight graduate students. Some tools were indicated more than once by individual participants. Figure 28 presents the results of this item showing the individual answers provided by each participant.

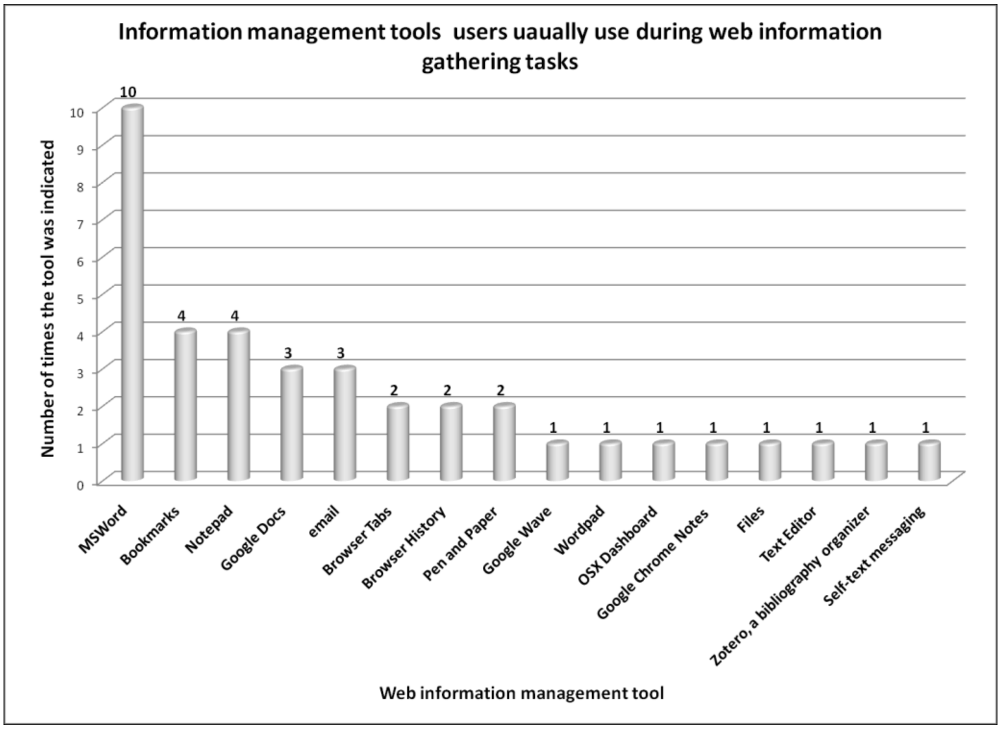


Figure 28. Tools users usually use to organize and manage web information.

The principal investigator met with a group of eight graduate students. The meeting was intended for grouping each set of similar tools under one category. The group agreed on the categories shown in Figure 29. The results indicated that text editors including—mainly—*MSWord* and *Notepad* in addition to browser features such as browser tabs were the most dominant organization tools for information management during web information gathering tasks according to the study participants. However, there was no significant difference between any two proportions of participants who indicated the use of any two different tools. After completing the study, the tools that were actually used by participants in the study were also analyzed.

Participants used several tools for organizing the task information. In addition to organizing by editing and formatting information in text editors, managing information took place in activities such as keeping and sending information. The tools used for organizing information for keeping and re-finding included bookmarking, files, and emails. Moreover, participants used browser tabs and windows (less frequently) to keep track of selected sources of information (pages) for comparison and decision making within the same session.

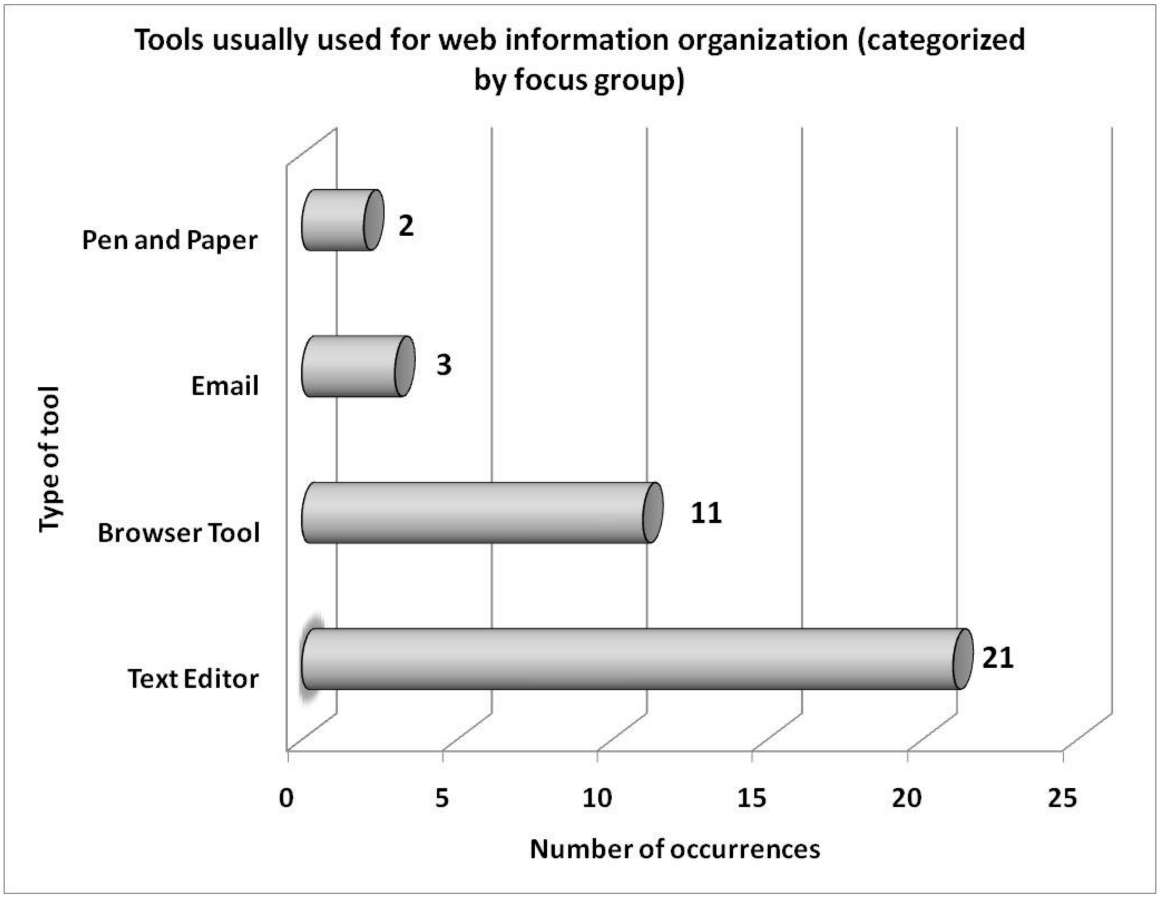


Figure 29. Categories of web information management tools.

Participants had to gather information required in each task over two sessions. In the second session, the task asked participants to re-find information that was located in the first part of the task (first session). This was done to stimulate users to organize information for re-finding and to investigate how users handle multiple sessions. The main goal was to reveal the level of effectiveness in current organization strategies and the difficulties associated with the organization process. The findings regarding this factor are discussed as part of the organization strategies in Section 5.3.6.

The results showed that over all the tasks performed in the study, participants used text editors 26 times (MSWord 13 times, Notepad 13 times) and used emails five times to organize and manage task information. There was a significant difference between the use of text editors and the use of emails in the study according to the *z*-test results (z =1.95, p<0.03). The results indicate that users rely heavily on editing to manage information gathered for the task. Interestingly although users used these tools, 95%

(19/20) indicated in the interview that none of the tools used for organizing the task information was sufficient. The reason they gave for using those tools was that they were the only ones available. As a result of using those tools, several difficulties during organizing and managing information were revealed as discussed in Section 5.3.7.

5.3.6.Organization Strategies

To organize and manage information for the task, users followed different strategies which included: keeping information using bookmarks, editing information in text files, editing information while composing email messages, keeping information in emails and local files, temporary keeping of pages on open browser tabs and windows, typing information located in web pages and search results summaries into text files and emails, copying and pasting information, and temporary storage of URLs in files and emails for comparisons and decision making. The following is a description of the strategies used during the task accomplishment process with the tools used for each strategy.

5.3.6.1.Keeping Information Using Bookmarks

Even though bookmarking is one of the most well-known features in the web browser for keeping links in order to re-find web pages, the users in the study did not use this feature much for the given information gathering tasks. As Figure 30 shows, 65% (13/20) of the users did not use bookmarks whatsoever. After removing the outlier, the average use of bookmarking in the study was 1.1 (SD=1.8). When questioned in the interview, users indicated that it was hard to re-find pages kept in a previous session and compare them to pages actively open in the browser during the current session. Users also indicated that completing the task required more than just keeping links to pages. It required access to parts of pages and other information such as comments entered by the user.

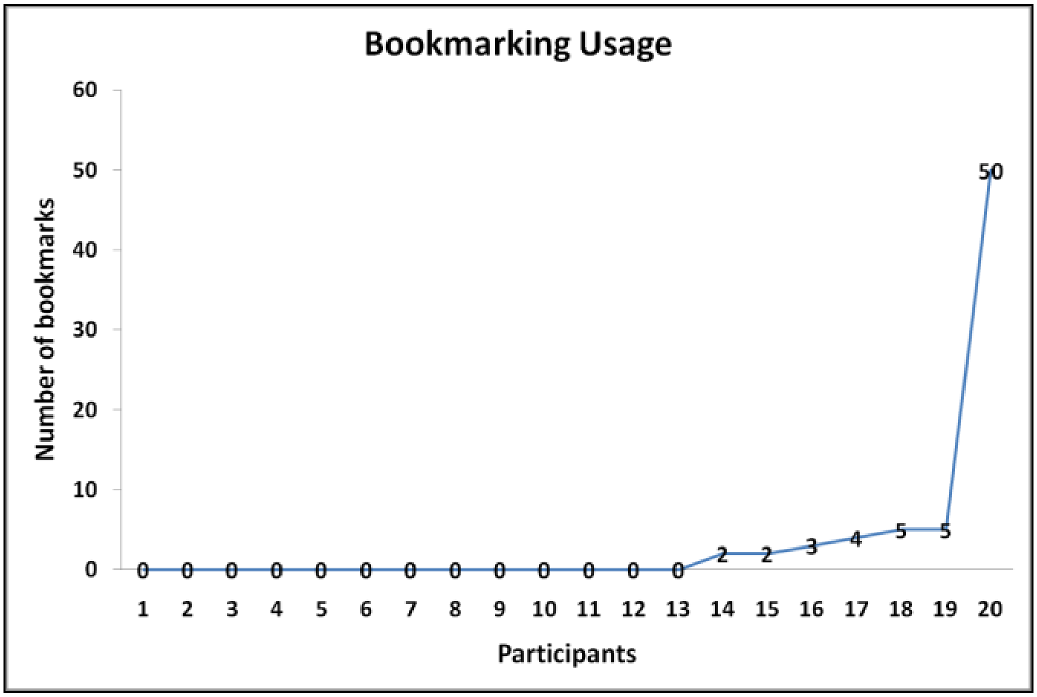


Figure 30. Use of bookmarks during the study.

5.3.6.2. Editing Files in Text Editors

Users in the study relied heavily on the use of text editors including MSWord and Notepad to organize the task information they had found. Of all uses of tools during the study, 55% was text editors' usage. Participants made use of text editors while browsing, searching, comparing, and selecting information from web pages.

All users in the study used text editors to edit and format information they found in web pages. More than half of the participants (12/20) extracted information from result hit summaries for editing and further formatting in text editors. The remaining users (8/20) copied information from web pages that were active (open) in the browser and pasted the information into text editors for further organization. In all cases, users reported that the process of accumulating, organizing, editing, and formatting the task information in text editors and email drafts was demanding. However, they indicated that using text files made it easier to manage and remember information than using the browser history or bookmarks for keeping track of web pages.

5.3.6.3. Emailing for Keeping and Later Re-finding Information

The use of emails was mainly for keeping and later re-finding information during the tasks. Only three participants in the study (as shown in Figure 29) used emails for either saving information for later re-finding or to explicitly satisfy the task requirements by collecting and sending the information in an email message. In this case, the email composing utility was basically used as a text editor. Users indicated that they used emails to allow them to find the information faster for later reuse in multi-session tasks due to the easiness in searching the content of email messages and drafts. Those users stated that using text editors and creating text files would result in difficulties in finding the information later since this process implies that some folder and file structure organization may become necessary.

5.3.6.4. Using Browser Tabs and Windows

To manage web pages during a session, the participants relied heavily on the use of browser tabs. On average, users opened 25 tabs during the entire study (SD=22.5) as shown in Table 5. Of the participants, 95% (19/20) used at least two browser tabs to keep track of open pages, to compare information, and to keep information for later use within the same session. They indicated that they used tabs rather than windows or bookmarks because they could see the information faster by switching among tabs, they were able to keep all pages open, and they were able to compare the information for decision making. Nonetheless, 40% of the participants (8/20) indicated that they lost track of pages those participants opened on browser tabs and they used only a few of the tabs they opened during the task. When asked, users indicated that this was the best available strategy offered by the current browsing model for organizing pages during a session.

5.3.6.5. Copying, Pasting and Typing Information for the Task

While most users used the copy-paste feature, the use of this feature had two variations. 40% of the participants (8/20) used copying and pasting information (as shown in Figure 31) from web pages into text files, emails, and so on. These users also frequently opened pages by clicking links to find information. The interviews revealed that it was convenient for those users to copy and paste the information directly from web pages. The remaining users (12/20) rarely used this feature. Instead, they opened more browser tabs to frequently submit search queries and extract or type into their word processor or email information from hit summaries provided by the search engines rather than from open web pages. Since hit summaries usually contain incomplete sentences, those users typed and edited the information into the text editors.

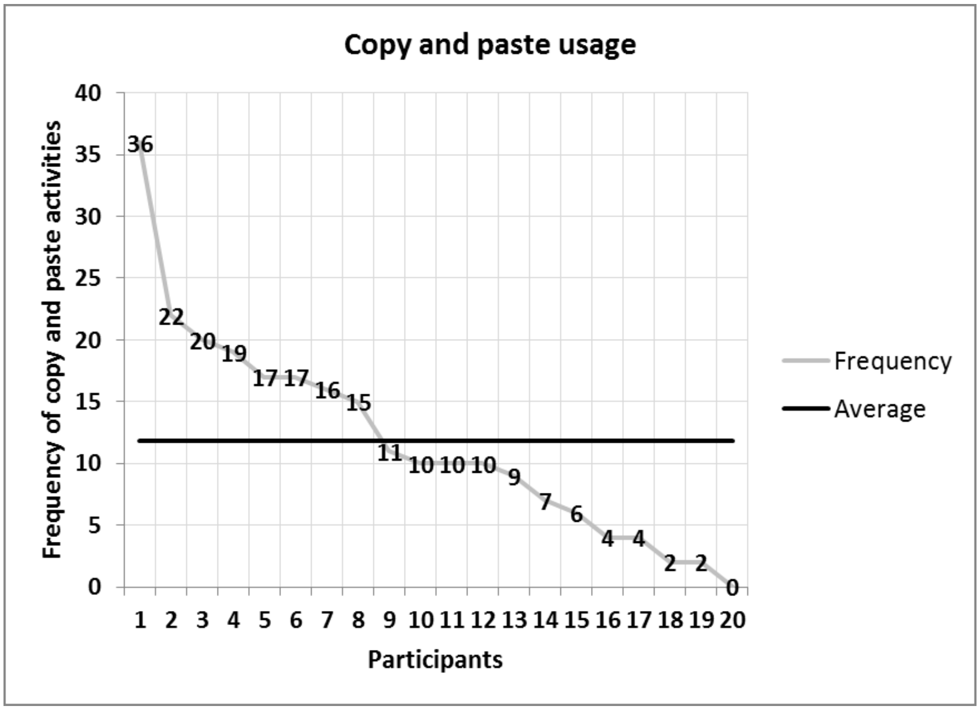


Figure 31. Copy and Paste usage during the study.

5.3.6.6. Keeping Links in Emails, Files, and Bookmarks for Re-finding

Users kept links along with information of interest in text files and email drafts. They reported a few main reasons for saving links, shown in Table 7. The most common reasons for keeping links were: to reference content already found (50%), to revisit a page (42%), and for comparisons and decision making (8%). Of the participants, 95% (19/20) stated in the interview that they preferred to keep links in files and emails rather than in bookmarks to have links related to a task in one place. Users also reported using those links as references to information, to remind themselves where to find more information for the task, and as sources for comparing information. One user indicated that she would like to have ‘automated referencing’ to embed links in the gathered information.

Table 7. Why users kept links.

Reasons for storing links in documents being edited	**Number of responses**
To reference information	20
To revisit later	17
To compare information	3
Total	**40**

The use of text files and email drafts for keeping track of web pages indicates the ineffectiveness of current re-finding techniques in web browsers—mainly bookmarking. Of the participants, 65% (13/20) did not use bookmarks. They used files and email drafts to keep track of links needed for re-finding and comparing information and web pages. They stated that it is easier to re-find pages of interest in files of known names and locations to the user than to use bookmarking. The heavy reliance on text files indicates that: first, editing while organizing information for the task is a substantial part of the gathering process; second, current re-finding techniques lack effectiveness for this type of task in which re-finding can happen not only over multiple sessions but also within the same session. Participants opened a large number of tabs, dealt with multiple windows simultaneously, and edited and managed information while trying to find information on the web.

5.3.7.Information Organizing and Management Related Difficulties

Prior to performing the tasks, users indicated that they usually encounter several kinds of difficulties while gathering and organize information on the web. Those difficulties and problems were concerned with: losing information due to switching among several tools, reliability and accuracy of information, finding useful summaries of information, organizing and managing information, information overload in web search results, re-formatting information copied from websites, losing links during long term tasks, and several other less significant issues. After performing the tasks in the study, users stated the information gathering problems they encountered during the task accomplishment process.

The problems encountered during the tasks of information gathering in the study included very specific points that were categorized by a group of eight graduate students. The categories involved organizing, storing, prioritizing, and filtering information as the most frequent problem encountered. This category represented problems with storing and organizing information, formatting information, filtering through information, copying and pasting information, editing information required in the task, switching among pages opened on browser tabs during the task, and prioritizing sources of information for selecting the task requirements. The main categories concerning information gathering difficulties in the study are shown in Table 8.

Table 8. Difficulties encountered during the tasks.

Type of difficulty	Number of times indicated
Organizing, storing, filtering, and prioritizing information	10
Difficulties with the web browser	5
Time difficulties	5
Difficulties with finding information	4
Not enough background knowledge about the task	2
Difficulties with having to deal with more than one session	1

To further understand the difficulties indicated during the study, users were further questioned during a short interview. The results show that users deal with the available tools on the web only because of their availability. One of the main concerns for participants in the study was their inability to edit and browse information simultaneously and that none of the available tools could satisfy all the task requirements. Participants indicated that even though they used many tabs to view, compare, and select information on the web, they lost track of pages opened on those tabs. In the study, participants opened an average of 25 browser tabs for the two tasks. Participants also stated that they kept URLs of web pages in files and email drafts for later re-finding and re-using because they had always found bookmarking to be ineffective. They indicated that they usually forget what they had bookmarked and for which task. Therefore, they saved links and information in files and email drafts created for the task at hand.

In addition to the information organization and management-related difficulties, participants encountered some issues with the web browser. Difficulties with the web browser included switching among open browser tabs and windows, having to read much text, and ineffectiveness with regard to the use of browser bookmarks as indicated by the participants in the questionnaires. Furthermore, users indicated that they had difficulties related to dealing with more than one session in terms of giving priorities to the information kept for later re-use and in terms of re-finding such information. The difficulties encountered during information gathering while users managed and organized information for the task show that the current web browser was not sufficient for the given information gathering tasks in the study. Users needed tools that allow for viewing multiple sources of information (pages) for comparison and decision making, allow for editing information altogether with other gathering activities, permit the user to preserve and keep different kinds of information including hyperlinks for later re-use, allow users to re-find information effectively and efficiently, permit the user to handle multiple sessions more effectively, and allow for effective organization of the task requirements.

Even though participants used browser tabs very frequently, they claimed to have lost track of open tabs. They also indicated that using multiple browser windows was even less effective while gathering information since it becomes hard to manage multiples windows open simultaneously. In addition, switching among different tools, which were needed for accumulating the tasks, made it harder to organize information. For example, one user needed the browser open with multiple tabs in an active state in addition to a calculator and a text editor. This user was also attempting to use an email draft to preserve links to pages for later reuse. When asked about the reasons for not using bookmarks, the user indicated that it would be easier to remember where the information had been kept if she used her email. Those strategies highlight the ineffectiveness of the tools currently available for web information gathering.

5.3.8.Other Activities and Tools

To organize information gathered for the task, participants performed additional activities and used certain tools with particular tasks; yet less frequently. For example, 10% (2/20) of the users used calculators to decide on what trip to choose for a task. Twenty percent (4/20) of the users used online services such as currency exchange calculators. One participant used the browser history to look up pages previously found of interest to that user. Among the less frequent activities during the tasks was storing complete pages temporarily while gathering further information. Other activities included using global bookmarking through Yahoo's delicious service[12]. This was used for looking up pages through searching global bookmarks but not for re-finding purposes.

[12] http://www.delicious.com/

5.4. Study Limitations

There are three main limitations to the user study conducted in this research. First, the study used simulated tasks. Those tasks do not reflect every possible information gathering task on the web even though the researchers took every opportunity to make the tasks as realistic as possible and the users' understanding of the tasks was very high. Second, all participants were computer science students. Those users do not reflect all web users. Finally, there are some features and browser add-ons that might have helped with the tasks and which were not involved in the version of the web browser used in the study.

5.5. Discussion

A user study was conducted, to investigate the current state of web information gathering tasks with respect to how web users organize and manage information during the task. The study was aimed at revealing tools users currently utilize to accomplish this type of task, difficulties they encounter during the task, and to recommend design principles for further research.

With regard to the tools used for organizing and managing web information, the results revealed that current browser features such as bookmarking were ineffective for re-finding as a part of information organization. In addition, the lack of the ability to edit web information while locating, comparing, and organizing such information for the task resulted in complaints from the users in the study. Users found it hard to perform those activities simultaneously since current tools did not allow for effective gathering and organization of web information. Moreover, the use of the browser tabs for keeping track of information during the gathering process resulted in losing information and ineffective comparisons for decision making regarding information selected for the task. Although browser tabs were the best available method for organizing active pages for comparison and gathering, they did not serve participates best as indicated in the interviews.

With respect to the strategies users adopt to accomplish information gathering tasks, there were several strategies some of which are worthy of further analysis. First, users tended to store links into text files instead of bookmarks or favorites. They even stored complete pages on the computer for later re-finding and re-using and to handle multiple sessions. In addition, users had to switch among several tools very frequently. They also had to switch among browser windows and tabs to keep track of pages and select, compare, and copy information from those pages. Participants also used their email 'compose' feature to store links and information for later re-finding. To organize information gathered for the task, users created folders and text files. Some of the information was stored in files which were used either temporarily or as containers of the final task results.

Interestingly, users indicated that they used browser tabs for organizing information for comparisons because this was the best available feature on the browser for this activity. They also had to switch between editing and browsing for organizing and gathering information due to the continuous need for the two activities for this type of task. Furthermore, 14 users (70%) indicated that they do not usually use bookmarks. They stated that they rarely bookmark or they bookmark and never go back to re-find the information they kept through bookmarking. These findings highlight several factors that need to be investigated in further studies. The results of the study provide the following insights:

- Bookmarking was ineffective for re-finding or handling multiple sessions.
- Keeping dispersed parts of the task information in files, emails, and saved pages was ineffective for handling multiple sessions.
- Users want the capability to edit and format web information during sessions.
- Browser tabs used for keeping track of information during a session resulted in losing information and were ineffective.
- Users had trouble managing information with the use of several tools for different activities simultaneously.
- Users adopted several strategies to gather their information using varied tools during the tasks. Those strategies involved:
    - Users stored links into text files instead of bookmarks or favorites. Most users (65%) indicated that they rarely, if ever, used bookmarks.
    - Users stored complete pages on their computer for later reuse.
    - Users created folders and text files to organize the information both for intermediary use and for the final results.

- Users used browser tabs for organizing information for comparisons because this was the best available feature on the browser for this activity.
- Users also had to switch between editing and browsing for organizing and gathering information due to the need for the two activities for this type of task.
- Users used their emails to store links and information for later re-finding.

From the findings of the study, several guidelines that would benefit the design of tools intended for web information gathering are proposed. The guidelines include:

- The design of such tools should allow users to search and browse for information while being able to extract and manage information for further editing and formatting.
- The design of information gathering tools should take into account the ability of the user to handle multiple sessions by keeping the task information integrated instead of different dispersed parts using different tools and strategies.
- Re-finding the task information should be considered by allowing the user to keep track of sources of information and references more effectively.

5.6. THE MODIFIED INFORMATION GATHERING MODEL

Based on the results of the study, the model established and modified as discussed in Chapters 3 and 4 was further considered for adjustments. From the data accumulated in this study, a separate subtask emerged. The new subtask is called handling multiple sessions. The subtask concerns how users stop and later restart an information gathering task over multiple sessions. This subtask involves shaping the task information to be saved and later relocated so that the user continues working on the same task in a subsequent session. Handling multiple sessions may involve aspects of keeping and re-finding information. In the current state of information gathering on the web, this subtask is performed following different strategies such as keeping links of pages, files of information, bookmarks, session marks, and so on. The modified model of information gathering, which involves the subtask of handling multiple sessions, is shown in Figure 32.

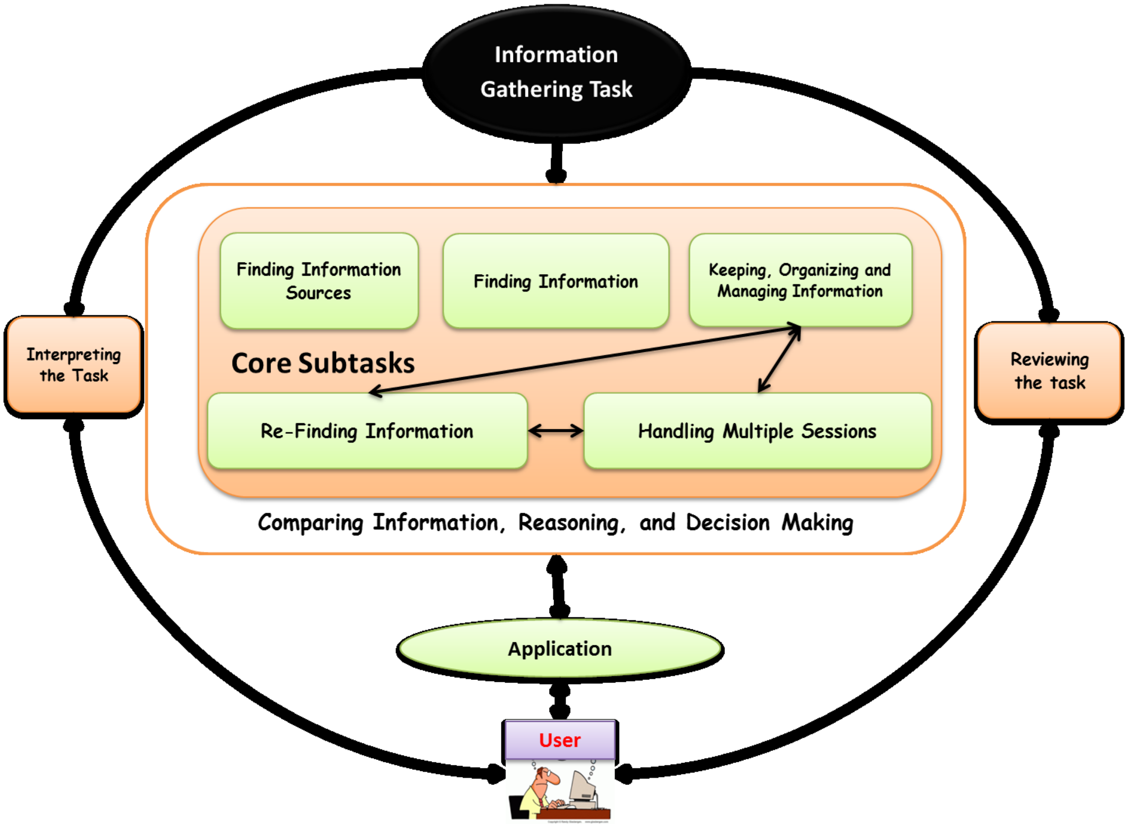

Information Gathering Task
Finding Information Sources
Finding Information
Keeping, Organizing and Managing Information
Core Subtasks
Re-Finding Information
Handling Multiple Sessions
Comparing Information, Reasoning, and Decision Making
Interpreting the Task
Reviewing the task
Application
User

Figure 32. The modified model of the information gathering task on the web.

5.7. SUMMARY

This chapter presented a user study in which the process of information gathering on the web was examined. The investigation was intended to reveal issues users have with each subtask involved in the overall task of information gathering on the web. The study showed that three subtasks require further investigations for improving how users accomplish those subtasks and the task of information gathering on the web as a whole. Users in the study had issues with re-finding the task information, handling multiple sessions, and managing and organizing the information for the task. Those issues will be considered for further investigations. Further consideration will be given to support these subtasks through designing and implementing features in web tools used for information gathering. The following chapter discusses those features and the results of the investigations.

CHAPTER 6 SUPPORTING INFORMATION GATHERING TASKS ON THE WEB

6.1. RESEARCH STUDY

In Chapter 5, the research investigated the task of information gathering on the web using the web browser in addition to other applications and tools required and needed for the task. The study was intended for developing recommendations for further work. The investigation identified: difficulties users have with information gathering tasks on the web, tools and features required for completing the task more effectively, and strategies users adopt to fulfill the task requirements. The main recommendations that were developed from the exploratory study described in Chapter 5 are:

1. Users should have the ability to re-find, information from sources located in past sessions more effectively.
2. Users should be able to handle multiple sessions more effectively. Users should have the ability to keep the task information integrated in one unit to preserve the task context and to be able to re-locate the task, not dispersed parts of it.
3. User should be able to manage and organize the task information more effectively. They should have the ability to search, browse, and edit information using one tool.

6.2. RESEARCH QUESTIONS

There are two research questions based on the recommendations from the previous studies. The research questions are:

1. Do the behavioural characteristics of users performing information gathering tasks on the web confirm the recommendations developed in the earlier studies? This question is answered by using:

   1.1. The data logged during the study.

   1.2. The user responses to issues regarding their behaviour while performing information gathering tasks.

2. What is the effectiveness of each of the following specific features:

2.1. Keeping track of references for web pages using thumbnails accumulated altogether in a reference tracking area compared to conventional methods such as copying and pasting links into text files in the case of the subtask of re-finding information.

2.2. Keeping and retrieving the task information integrated compared to keeping and later retrieving parts of the task using conventional strategies such as saving pages, saving information in files, bookmarking, and so forth in the case of the subtask of handling multiple sessions.

2.3. Using one application for searching, browsing, and editing compared to the use of multiple applications (i.e. the web browser and a text editor) in the case of the subtask of managing and organizing information.

The answer to these questions will be based on the user activities and their questionnaire responses. The numbers of user activities logged for each subtask provide indications of the frequency of the subtask and the effect of each feature implemented. The users' answers to the questionnaires identify issues related to each subtask from the point of view of the user. Analyzing the agreement/disagreement of those answers with the data logged in the study helps with answering the research questions.

6.3. Study Prototype: Design and Implementation

A prototype interface called WIGI (**W**eb **I**nformation **G**athering **I**nterface) was designed and implemented to investigate design features concerned with the three recommendations listed above. WIGI, shown in Figure 33, was built using JavaScript, ActiveX components, and HTML. WIGI consists of three main parts (areas). The following are the recommendations from the previous study associated with features implemented in WIGI.

1- Re-finding Information, the Reference Tracking Area.

a. Users can keep track of every URL clicked in the search results.

b. They can click each URL during any session within the same task to open the corresponding page.

    c. Links clicked from the search hit list are captured and shown to the user associated with the thumbnail and title of the page.

2- Handling Multiple Sessions, the Control Bar.

    a. Users can save the session information including: the tracked links, the information collected in the editor, and the links embedded as references as one integrated unit representing the task.

    b. Users can restart the task in subsequent sessions and have the information gathered in previous sessions retrieved as one integrated unit.

3- Managing and Organizing the Task Information.

    a. The Embedded Editor

        i. Users can drag and drop information from web pages into the editor.

        ii. They can add their input to the task using the editor.

        iii. Users can format the information in the editor as required in the task.

        iv. They can embed references into the information gathered in the editor.

    b. The Browsing Area

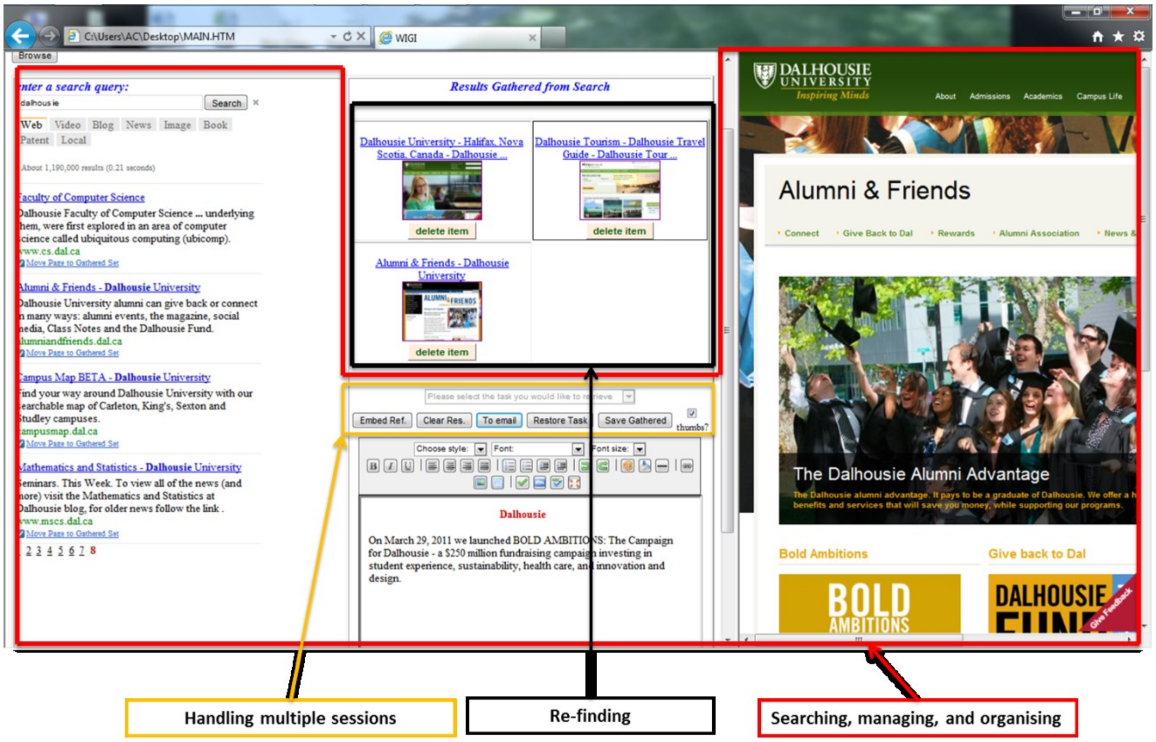
        i. Users can browse search results (pages) and typed in URLs on the same display along with editing, searching, and reference tracking.

    c. The Search Area

        i. Users can search the web for information using a search engine.

        ii. They can track every search hit clicked to appear in the reference tracking area.

        iii. They can browse search hits on one display along with the list of hits being viewed.



About 1,190,000 results (0.21 seconds)

Faculty of Computer Science
Dalhousie Faculty of Computer Science ... underlying hem, were first explored in an area of computer cience called ubiquitous computing (ubicomp).
www.cs.dal.ca
Move Page to Gathered Set

Alumni & Friends - Dalhousie University
Dalhousie University alumni can give back or connect n many ways: alumni events, the magazine, social nedia, Class Notes and the Dalhousie Fund.
alumniandfriends.dal.ca
Move Page to Gathered Set

Campus Map BETA - Dalhousie University
Find your way around Dalhousie University with our earchable map of Carleton, King's, Sexton and Studley campuses.
campusmap.dal.ca
Move Page to Gathered Set

Mathematics and Statistics - Dalhousie University
Seminars. This Week. To view all of the news (and nore) visit the Mathematics and Statistics at Dalhousie blog, for older news follow the link .
www.mscs.dal.ca
Move Page to Gathered Set

1 2 3 4 5 6 7 8

Dalhousie

On March 29, 2011 we launched BOLD AMBITIONS: The Campaign for Dalhousie - a $250 million fundraising campaign investing in student experience, sustainability, health care, and innovation and design.

The Dalhousie alumni advantage. It pays to be a graduate of Dalhousie. We offer a h benefits and services that will save you money, while supporting our programs.

Figure 33. WIGI's interface.

6.3.1.Re-finding Information: The Reference Tracking Area

The reference tracking area was implemented to assist users to re-find information. This area includes references opened by the user from the search result hits. The user can click a special link associated with every search hit to do two things. First, the link opens in the browsing area. Second, a thumbnail of the page clicked appears immediately on the reference tracking area associated with the title and URL of that page. This area is located at the top of the middle part of the display as shown in Figure 33. The references tracked are shown in a tabular format on the display. The user can click on any thumbnail at any time to open the corresponding page in the browsing area as part of the re-finding process. References are kept along with the task information for subsequent sessions while working on the same task.

6.3.2.Handling Multiple Sessions: The Control Bar

The control bar contains features intended to assist users with handling multiple sessions. It lies between the reference tracking area and the embedded editor. The control bar allows the user to save the current state of the task as one integrated unit of information. This information includes the references accumulated in the reference tracking area and the information collected in the embedded editor. The bar allows the user to embed references into the editing area to be associated with information being gathered. It also allows the user to send the information collected in the embedded editor to their email at any time. In addition, the control bar allows the user to restart a task that has been saved in a previous session.

6.3.3.Managing and Organizing Information

6.3.3.1. The Embedded Editor

The embedded editor allows users to manage and organize information they collect from web pages without switching between the web browser and other applications (i.e. editors or emails). The embedded editor appears in the middle of the display under the reference tracking area (Figure 33). Its position was selected to allow the user to drag and drop information from both web pages open in the browsing area and result hit summaries. The user can change the size of the editing area in the editor as desired. The embedded editor is fully loaded with editing and formatting features. It allows the user to drag and drop objects (text, images, links … etc.) from web pages directly into the editor while keeping the original format of the objects.

6.3.3.2. The Browsing Area

This part appears on the left side of the display as shown in Figure 33. The browsing area is a frame on the display that allows the user to see the content of each web page the user clicks from either the search hits in the search area or the thumbnails in the reference tracking area. Combining the browsing area along with the search and editing areas on one display was intended to minimize the need for switching among applications. The user can resize the browsing area according to the dimensions of the page being viewed. Having the browsing area along with the search area and the embedded editor was meant to assist users with managing and organizing the task information. It also helps the user to re-visit pages from the current and past sessions without losing the task context due to switching between browser windows or tabs.

6.3.3.3. The Search Area

In the search area (shown in Figure 33), users could send search queries (powered by Google) and receive search result hits. When the user clicks a search result hit, the page opens in the browsing view on WIGI by default unless the user forces it to open in a new tab or window. A thumbnail of the clicked page appears in the reference tracking area. The default number of search results shown to the user for each query is limited by the space available. A tabbed search bar; that allows searching for images, maps, and so forth; is available to the user at all times.

6.4. STUDY POPULATION

Thirty participants were recruited for the study. All of the participants were computer science students from Dalhousie University. Of the participants, 15 users were males and 15 were females. Fifteen participants were graduate students while the remaining were undergraduate students. Participants in the study were between the age of 18 and 30. Attrition did not occur at any time in the study.

6.5. STUDY DESIGN

The design of the study was complete factorial and counterbalanced. Four tasks were used in the study (the same tasks executed in Chapter 5). Every task had the same chance of being used in the study. The order of distributing the tasks over the tools (WIGI or browser) and participants was random. Every participant performed a total of two tasks with one task (divided into two parts) executed on WIGI and one task (also divided into two parts) on the ordinary browser. Both the browser and WIGI had the same chances of being used first. The browser used in the study was Internet Explorer (version 9). This browser was selected due to the need for using ActiveX components. The study had four conditions: two processes (browser + WIGI) and two tasks.

6.6. STUDY TASKS

For this study, the same tasks described in Chapter 5 were used. The tasks were created using guidelines from previous work (Kules and Capra, 2008) and further examined using a focus group as illustrated in Chapter 5.

6.7. STUDY METHODOLOGY

Each participant was randomly assigned two of the four tasks. Each task consisted of two parts each of which was performed during one of the two sessions of the study. On the first day of the study, each participant signed the consent form after being introduced to the study and after explaining the participant's role in the study. Then, the participant was given a short training session on WIGI (five to ten minutes). The participant then completed an online pre-study questionnaire shown in Appendix D. After completing the questionnaire, the participant performed the first part of the first task on either WIGI or the browser. Then, the participant was given the first part of the second task to complete on the tool (WIGI or browser) the participant did not use for the first task.

On the second day, each participant returned to complete the second session of the study. First, the participant completed the second part of the first task on the same tool (WIGI or browser) they used for the first part of the first task. Then the participant completed a post-task questionnaire for the first task. Then, the participant completed the second part of the second task on the same tool they used for the first part of that task which they completed in the first session. Afterwards, the participant completed a post-task questionnaire for the second task. Then, the participant was interviewed shortly for answering questions related to the way the participant completed the study with regard to why certain tools and strategies were used. The study procedure is depicted in Figure 34.

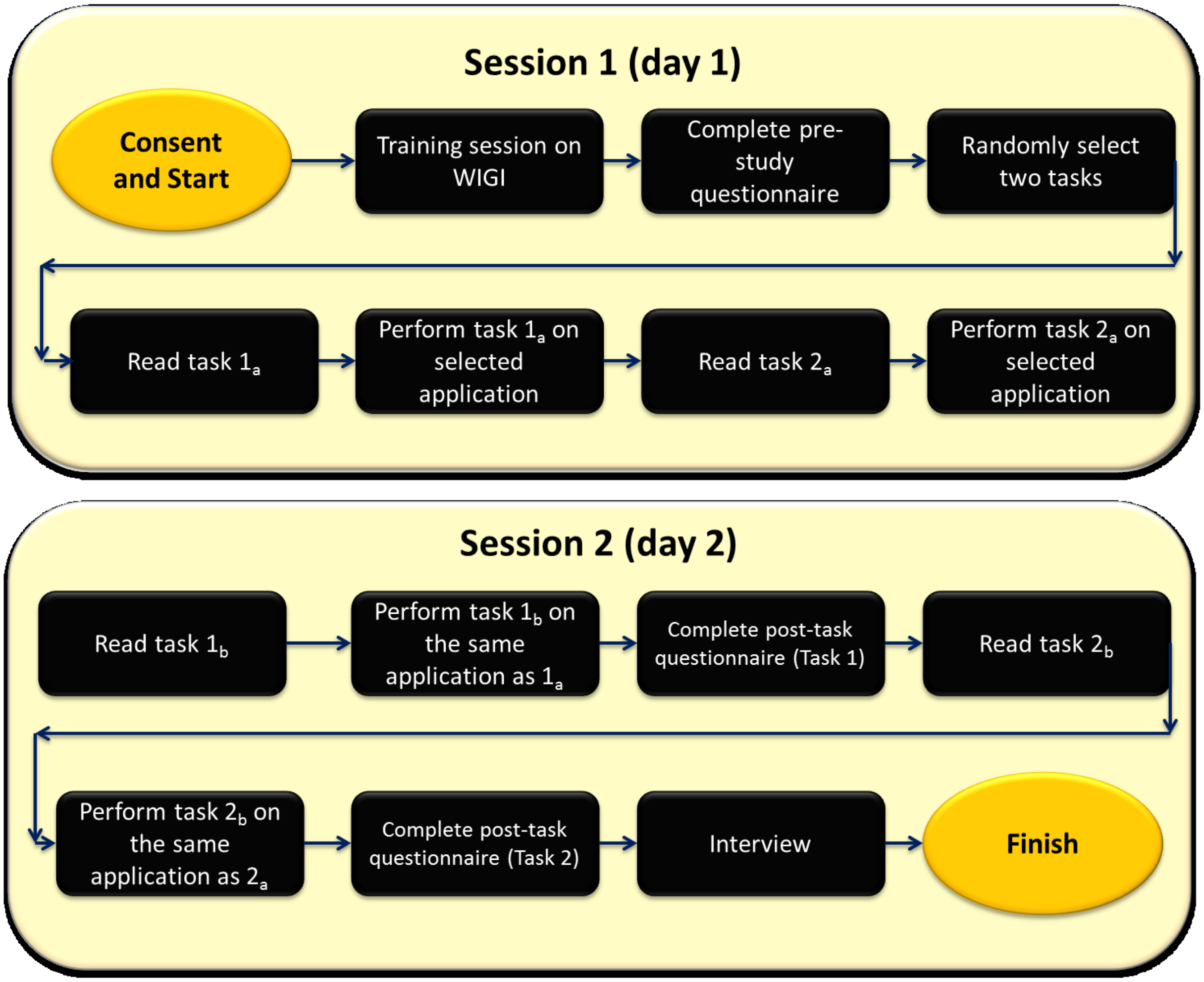

Session 1 (day 1)
Consent and Start
Training session on WIGI
Complete pre-study questionnaire
Randomly select two tasks
Read task 1_a
Perform task 1_a on selected application
Read task 2_a
Perform task 2_a on selected application
Session 2 (day 2)
Read task 1_b
Perform task 1_b on the same application as 1_a
Complete post-task questionnaire (Task 1)
Read task 2_b
Perform task 2_b on the same application as 2_a
Complete post-task questionnaire (Task 2)
Interview
Finish

Figure 34. Study procedure.

6.8. Study Results

The study data came from two sources: the log file of activities performed in the study and the questionnaires. Overall, 5436 activities were logged during the study. Of the activities, 2539 activities were recorded while using the browser and 2897 activities were recorded while using WIGI. The activities logged in the study were chosen prior to the study and are shown in Table 9. Every activity belongs to a subtask in the information gathering task. Those activities were selected to reveal the differences in the way users performed the tasks using the prototype features and using the conventional browser.

Table 9. Activities recorded in the study.

Category	User ID	Session ID	Task ID	System, ➢ A: WIGI ➢ B: Browser	Time	Current URL	Activity	Field 1	Field 2
Re-finding				A			Re-Open Task	Task label	
				A, B			Click Link in A-Editor	URL	
				B			Click Link in File/Email	URL	
				B			Re-open session	URLs	
Handling Multiple Sessions				B			Create Bookmark	URL	
				B			Save Page	URL	
				B			Create file		
				B			Save session	URLs	
				A			Save Task	Task label	
Managing Information				B			Open new tab	Blank/URL	
				A, B			Copy	Link/Other	Snippet/Page
				A, B			Paste	A-Editor/Other	
				A, B			Edit		
				A			Move	Snippet/Page	WIGI-Editor/Other
				A, B			Email		
				A, B			Format	Operation	WIGI-Editor/Other
				A			Embed Reference	URL	WIGI-Editor
Finding Information				A , B			Query	Keywords	Page / Search Engine / browser find
				A , B			Typed in URL	URL	
				A, B			Auto-complete URL	URL	
				A , B			Link clicked on menu	URL	
				A, B			Link clicked on page	URL	
				A, B			Link clicked on hit list	URL	
				A , B			Open bookmark		

The post-task questionnaires in the study were used to capture the user perceptions such as the perceived level of completeness of the task by the user, the perceived level of confidence in the task results, and other comments related to the tools used in the task. The pre-study questionnaires were used to capture more generic data such as the tools users usually use for information gathering on the web, the difficulties they often encounter, and other demographic data. The pre-study questionnaire can be found in Appendix D.

6.8.1.Pre-study Questionnaire Data

The pre-study questionnaire involved collecting data regarding the age of the user, their experience with web information gathering, the tools they usually use, and the difficulties they encounter while gathering information on the web. All users were under the age of 31 and the study had equivalent numbers of both genders (15 males and 15 females). All users (30) were regular web information gatherers. The tools users indicated using for information gathering tasks are shown in Table 10. The difficulties and issues the users reported with information gathering on the web are shown in Table 11.

Table 10. Tools and applications users usually use.

Tool or Application		Responses	# Users
Web Browser		100.00%	30
Text editor		86.70%	26
Local bookmarking		50.00%	15
Online bookmarking		23.30%	7
Session saving		23.30%	7
Other:			
	Email	6.00%	2
	Query Saving	3.00%	1
	Paper	3.00%	1
	Plug-ins	3.00%	1

As shown in Table 10, the web browser, the text editor, and local bookmarking were indicated as the most frequently used tools for information gathering. Other features and tools such as online bookmarking, session saving, and emails were indicated on very few occasions. Even though these results reflect what the users believe they use in usual, the study was expected to reveal findings that could differ.

Table 11. Issues users reported having with information gathering tasks on the web.

Related Subtask	Issue	# Participants	% Responses
Handling Multiple Sessions	saving pages	19	63.30%
	saving information together	13	43.30%
	re-locating a task on which the user worked in previous sessions	11	36.70%
	saving sessions	10	33.30%
Re-finding Information	creating bookmarks	12	40.00%
	retrieving bookmarks	12	40.00%
Managing and Organizing Information	editing along with browsing for information	9	30.00%
	searching along with browsing for information	8	26.70%
	searching and managing information for a task	5	16.70%
	saving open tabs together	5	16.70%

As shown in Table11, the issues indicated by high percentages of user responses are those related to handling multiple sessions and re-finding information during information gathering tasks on the web. Managing and organizing information had fewer issues as shown in the user responses. Interestingly, these three main issues confirm the recommendations from the previous study. Consequently, those recommendations are consistent with the user responses which answers the second part of the first research question (Section 6.2).

Handling multiple sessions has issues including saving pages and sessions, saving the information of the task integrated, and re-locating the task instead of parts of the task. With regard to re-finding information, users were concerned with creating and re-opening bookmarks as the main issue encountered for the purpose of re-finding. Lastly, with regard to managing and organizing information during gathering, users were concerned with editing and browsing simultaneously as well as editing and searching. To a lesser extent, users seem to have had issues with dealing with open browser tabs. The following section discusses the study results in the light of the recommendations illustrated in Chapter 5 and highlighted above.

6.8.2.Re-finding Information

The study logged activities related to re-finding information in the case of the browser and WIGI as shown in Table 5. The re-finding activity used in the analysis of the study data was revisiting references (links to web pages) to information accessed in the first session.

On WIGI, users made 185 (3.4% of the total activities) re-finding activities while they made only nine re-finding activities on the browser (0.16% of the total activities). The difference between the two cases was significant according to ANOVA (F *(1, 58) =14.15, p<0.0005*). The results for the re-finding activities in the study are shown in Table 12. Those results show that users who performed frequent re-finding activities on WIGI seldom revisited any links when using the browser. Consequently, the difference cannot be attributed to individual preferences. The average re-finding activities on WIGI was 6.17 (SD=8.51) and the average on the browser was 0.30 (SD=0.79).

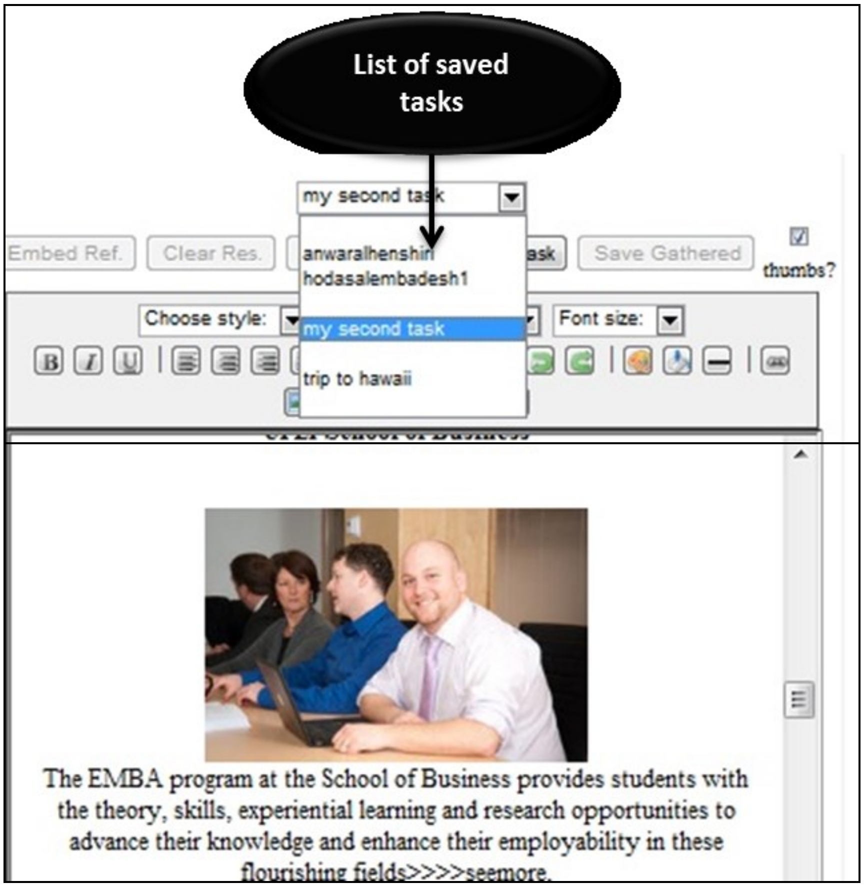
When asked, participants indicated that they did not attempt to re-visit links they kept in separate text files because they had to copy the links from the files and emails and paste them in the browser address line which would involve switching between two applications in addition to copying and pasting. On the other hand, WIGI had the links tracked and accumulated in the reference tracking area associated with thumbnails that were visible and clickable at any time. Thus, it was easier to revisit a link to see its content in the browsing area.

Table 12. Revisiting for re-finding activities.

Participants	**Re-finding (Link Revisited) WIGI**	**Re-finding (Link Revisited) Browser**
P29	42	0
P22	19	0
P12	17	0
P26	14	0
P24	13	0
P13	11	1
P8	7	0
P16	7	0
P21	7	0
P17	5	1
P9	5	0
P23	5	3
P4	5	0
P15	4	0
P30	4	0
P10	4	3
P28	4	0
P18	3	0
P3	3	0
P5	2	0
P27	2	0
P2	1	0
P11	1	0
P25	0	1
P7	0	0
P6	0	0
P20	0	0
P19	0	0
P14	0	0
P1	0	0
Total	**185**	**9**
Mean	**6.17**	**0.30**
SD	**8.51**	**0.79**

6.8.3. Handling Multiple Sessions

To save a task for subsequent sessions while using WIGI, all users used the *save gathered* feature built in the control bar. This feature allowed users to keep the task information integrated in one unit permitting them to restart the task in later sessions. Figure 35 shows the list of saved tasks and how the user selects from the list to restart a task. Users kept the information they had collected and organized in the editor and the references they accessed during the first session which were accumulated in the reference tracking area. During the second session, restarting the task required the use of the *retrieve task* feature, which is also built in WIGI. None of the users used any files or emails to handle multiple sessions with WIGI.



The EMBA program at the School of Business provides students with the theory, skills, experiential learning and research opportunities to advance their knowledge and enhance their employability in these flourishing fields>>>>seemore.

Figure 35. Handling multiple sessions.

On the browser, users used four different strategies to handle multiple sessions. Twenty six participants (26/30) created text files (using either MSWord or Notepad) to keep the task information and restart the task in the subsequent session. Four users (4/30) created 15 bookmarks. However, the same users reopened the bookmarks they created only 10 times. Four users (4/30) created email drafts to keep the information for subsequent sessions. Two users (2/30) saved complete pages to be used in the second sessions. Interestingly, neither of those two users re-opened the pages they saved.

The difference between the number of users who used the *save gathered* feature in WIGI (30/30) and the number of users who used text files to keep the task information in the case of the browser (26/30) was significant (z-test, *z=2.15, p<0.04*).

6.8.4. Organizing and Managing Information

To manage and organize the task information, users followed different strategies on each tool (WIGI or browser). All users indicated that they understood the tasks and had no problems with the descriptions of the tasks. To capture how users managed and organized the task information, different activities were logged and analyzed as shown in Table 5. These activities are of two types: activities performed on both WIGI and the browser and activities performed on either WIGI or the browser. Activities of the first type included: formatting, typing, copying and pasting, result hits clicking, menu and page link clicking, and reference embedding/reference copying and pasting. In the case of the browser, some of these activities required the use of other applications such as emails and text editors. The activities performed only on the browser included: creating bookmarks, opening bookmarks, creating files, opening files, closing tabs, creating email messages, and opening email messages. Activities that were available only on WIGI included using and removing thumbnails.

6.8.4.1. Copying and Pasting Information

One important activity related to managing and organizing the task information is copying and pasting information during the task. The study logged copying and pasting information from web pages into the information pool (e.g. editor) where the user collected the task requirements. Other copying and pasting activities such as between files (11 times) or emails (5 times) or within the embedded editor (13 times) on WIGI happened infrequently and were ignored in the analysis. Moreover, since some copying activities were not completed by the user (no subsequent pasting took place), the study considered the successful pasting activities only. The study recorded 330 pasting activities in total (6.01% of the total activities).

The copy and paste data are shown in Table 13.On WIGI, users performed more copying and pasting than they did on the browser (214 times vs. 116 times). The ANOVA test results (*F(1,58)= 5.6, p<0.02*) show a significant difference between the number of pasting activities on WIGI and the number of pasting activities in the case of using the browser. It is worth noting that in the case of using the browser, the user had to copy information from web pages into a separate application such a text editor or an email draft. The data also indicate that the difference in use between the two cases may be attributed to individual differences among participants. Users with more copying and pasting while using WIGI tended to copy and paste more when using the browser.

Table 13. Copying and pasting activities.

Participants	**Pasting (WIGI)**	**Pasting (Browser)**
P25	22	3
P5	20	16
P7	14	10
P12	14	3
P27	13	3
P6	11	11
P20	10	8
P19	10	8
P14	10	0
P18	10	0
P15	10	1
P1	9	0
P17	8	9
P2	8	6
P9	7	7
P13	7	11
P29	6	2
P22	6	13
P26	5	2
P30	5	0
P8	3	0
P16	3	0
P24	1	0
P23	1	0
P11	1	0
P4	0	0
P21	0	2
P10	0	1
P3	0	0
P28	0	0
Total	**214**	**116**
Mean	**7.13**	**3.87**
SD	**5.84**	**4.75**

6.8.4.2. Typing Information

While gathering information, not only do users copy information from web sources(pages) but they may provide their own input to the task or perform re-phrasing such as when they write a report or a survey article. Users may type information along with the information they find on pages such as to make their own conclusions. Every time the user hit letter or number keys on the keyboard, the activity was considered typing. A typing activity ended with the use of the mouse or a control key such as the *carriage return* or the *tab* keys. On WIGI, users did not type information as frequently as they did in the case of using the browser. The ANOVA test results (*F (1, 58) =4.40, p<0.05*) showed a significant difference between the number of typing activities on WIGI and those performed on the browser. In addition, the data shown in Table 14 for the typing activities indicate that the difference between the use of WIGI and the browser with respect to typing activities cannot be attributed to individual differences. Users behaved differently using the two tools indicating that the difference may be attributed to the tool used.

In further analysis, the correlation between pasting and typing information in the case of WIGI and the browser was computed. On WIGI, the correlation between typing information and pasting information was not significant (Pearson Product Moment, *r= -.29, p<0.12*). On the browser, the correlation was not significant (*r = 0.36, p<0.06*). As a result, it is impossible to use the correlation test results to explain the relationship between typing and pasting behaviours on either WIGI or the browser.

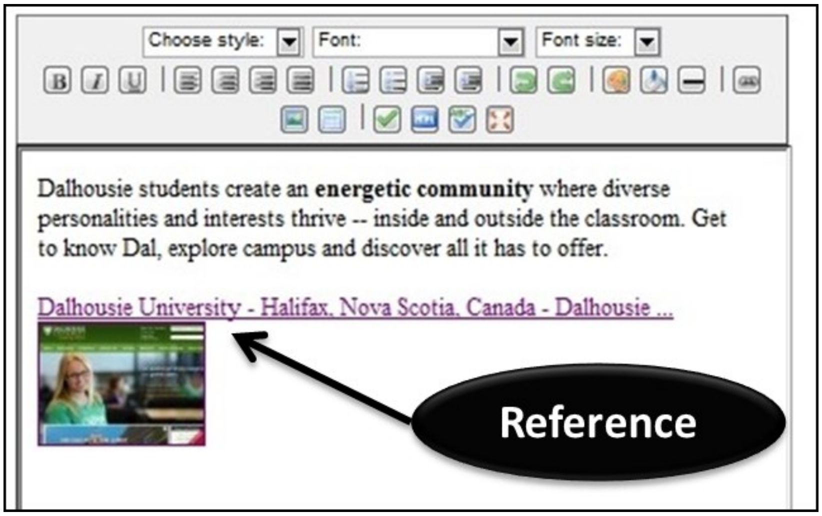
Table 14. Typing activities.

Participants	Typing (WIGI)	Typing (Browser)
P20	20	25
P19	19	22
P4	15	5
P21	14	13
P26	9	16
P8	9	16
P24	9	10
P14	8	7
P9	8	16
P13	8	18
P23	8	7
P10	8	4
P3	8	16
P7	7	2
P30	6	2
P11	6	6
P27	4	16
P6	4	17
P29	4	3
P28	4	6
P25	3	7
P12	3	8
P18	3	5
P15	3	5
P16	3	0
P5	2	6
P17	2	7
P22	1	12
P1	0	3
P2	0	12
Total	**198**	**292**
Mean	**6.6**	**9.73**
SD	**5.06**	**6.44**

6.8.4.3. Embedding References to Manage Information

In this context, embedding references is an activity intended for keeping links to web pages as part of the managing and organizing subtask to produce the final form of the task information. The study recorded a total of 410 referencing activities (7.54% of the total activities). On WIGI, users used the *embed reference* feature provided on the Control Bar. While a page is open in the browsing area, the user could embed its link in the editor. The link then would appear where the cursor was located in the editor associated with the title of the page (the title is associated with the URL), see Figure 36. The reference could be embedded with or without the thumbnail of the page. Thumbnails kept along with URLs and titles were intended to provide the user with visual clues about the page. There was no significant difference (ANOVA [*F (1, 58) =2.6, p<0.12*]) between the number of cases where thumbnails were used (83 times) and the cases where references were embedded without thumbnails (131 times).

On the browser, users copied and pasted links into text files or emails to keep track of their references. The data are shown in Table 15. The ANOVA test results showed no significant difference between the number of times users embedded references using WIGI and the number of times users copied and pasted URLs for referencing while using the browser (*F (1, 58) = 0.67, p<0.42*).



Dalhousie students create an **energetic community** where diverse personalities and interests thrive -- inside and outside the classroom. Get to know Dal, explore campus and discover all it has to offer.

Dalhousie University - Halifax, Nova Scotia, Canada - Dalhousie ...

Figure 36. Reference embedded in the editing area.

Table 15. Referencing activities.

Participants	Reference Embedded (WIGI)	Link Pasted (Browser)
P14	12	11
P23	8	11
P26	7	10
P9	8	10
P28	7	10
P20	4	9
P27	8	9
P25	9	9
P5	10	9
P22	6	9
P19	4	8
P10	7	8
P18	4	8
P8	5	7
P13	5	7
P3	9	7
P6	10	7
P4	11	6
P7	11	6
P21	7	5
P24	11	5
P12	9	5
P17	3	5
P11	4	4
P15	4	4
P29	6	3
P1	6	3
P2	6	1
P30	5	0
P16	8	0
Total	**214**	**196**
Mean	**7.13**	**6.53**
SD	**2.52**	**3.09**

6.8.4.4. Formatting Information

Formatting information collected for the task of information gathering includes using headings for the gathered text, changing fonts and colors, moving objects within the gathered information (within a file, an email draft… etc.), and resizing objects such as images. These are examples of formatting activities logged during the gathering process.

Table 16. Formatting activities.

Participants	Formatting (WIGI)	Formatting (Browser)
P27	79	6
P1	52	0
P18	38	0
P25	37	0
P6	36	6
P9	30	9
P20	27	7
P19	27	7
P12	26	0
P14	21	0
P15	20	0
P29	15	0
P7	13	0
P22	10	0
P26	10	3
P5	9	1
P30	9	0
P4	8	0
P10	8	0
P23	7	0
P21	7	1
P3	7	0
P8	6	0
P2	5	4
P11	4	0
P16	3	0
P13	2	7
P24	2	0
P28	2	0
P17	1	0
Total	**521**	**51**
Mean	**17.37**	**1.7**
SD	**17.63**	**2.88**

On WIGI, users used the formatting features in the embedded editor built in WIGI to format the task information. The embedded editor provided several formatting features such as font formatting, tables, and image formatting. Users performed an average of 17.37 (SD=17.63) formatting activities while using WIGI. The number of formatting activities on WIGI was 521 (9.58% of the total activities). On the browser, using other applications, users performed an average of 1.70 (SD=2.87) formatting activities during the study. The number of formatting activities was 51 (0.93% of the total activities) while using the browser. The data are shown in Table 16. The difference between the number of formatting activities on WIGI and the browser was statistically significant according to ANOVA (*F (1, 58) =23.08, p<0.002*).

The correlation between the activities of typing and formatting was considered. Measuring the correlation was intended to explain whether or not users who typed in more information (as opposed to pasting) did more formatting. The results of the Pearson Product Moment correlation test showed that in the case of WIGI, the correlation was weak with inverse relation (r = - *0.11, p=0.53*), i.e. not significant. In the case of the browser, the correlation was strong and positive (*r = 0.77, p<0.001*). This indicates that in the case of WIGI, users who did not type much in the first place also did not perform much formatting since WIGI allowed them to copy and paste the information with its original formatting (as later explained by the users). In the case of the browser, however, the correlation explains that as users did more typing, they followed with more formatting.

6.8.4.5. Finding Information

Finding information is a fundamental subtask of the information gathering task. The finding activities recorded in the study included: search queries submitted to search engines, search queries submitted on web pages, links clicked on web pages, result hits clicked, and the use of find-on-page feature in the browser. These activities allowed users to find information sources and to find information on the located sources. The total number of finding activities was 955 (17.57% of the total activities). The total number of finding activities on WIGI was 565 while the total number of finding activities on the browser was 390.

On WIGI, users submitted 250 queries to Google (the underlying search engine) to locate information sources (web pages). None of the participants typed in URLs to start searching for information. Similarly, participants submitted 251 queries to search engines in the case of using the browser. Google was the dominant search engine used in the case of the browser. There was no difference between the browser and WIGI with regard to the numbers of search queries submitted to search engines. The number of querying activities represented nine percent of the overall activities in the study. This indicates the complexity of the task of information gathering which requires more than submitting search queries for finding information independent of the tools used.

Table 17. Finding activities.

Participants	**On-Page Link Clicked (WIGI)**	**On-Page Link Clicked (Browser)**	**Search Result Hit Clicked (WIGI)**	**Search Result Hit Clicked (Browser)**	**Find on Page (WIGI)**	**Find on Page (Browser)**
P20	16	0	29	1	2	0
P19	16	9	28	2	0	2
P21	12	0	22	0	0	0
P18	10	12	20	7	0	0
P2	9	21	19	2	0	0
P23	7	18	19	13	0	0
P13	6	4	18	1	0	0
P26	6	5	17	12	1	0
P28	6	7	17	22	0	0
P12	5	4	17	9	0	0
P10	5	3	16	4	0	0
P5	4	3	16	3	0	0
P14	3	6	16	0	0	0
P11	3	8	15	0	0	0
P4	3	3	15	2	1	0
P3	3	15	13	11	0	0
P27	2	5	13	5	0	0
P17	2	0	13	3	0	0
P9	2	8	13	12	0	0
P8	2	7	11	5	0	0
P25	1	2	11	5	0	2
P6	1	23	10	13	0	0
P30	1	3	10	5	0	0
P16	1	0	10	13	0	0
P24	1	1	9	0	0	11
P7	0	0	9	5	0	0
P15	0	3	7	4	0	0
P1	0	0	7	3	0	0
P29	0	2	7	5	2	2
P22	0	12	5	12	0	10
Total	**127**	**184**	**432**	**179**	**6**	**27**
Mean	**4.23**	**6.13**	**14.40**	**5.97**	**0.20**	**0.90**
SD	***4.47***	***6.32***	***5.79***	***5.35***	**0.55**	**2.68**

Users of WIGI did not submit any search queries using the search box provided on some web pages. On the browser, six users (6/30) submitted a total of 22 queries to find information on web pages. Even though there is a difference between the two cases, the number of search queries submitted while using the browser was very small. With respect to the number of links clicked to navigate through websites while searching for information, the data are shown in Table 17. The results indicate no significant difference between the links clicked on pages in the case of WIGI and the number of links clicked in the case of the browser (ANOVA*, F(1, 58) = 1.8, p<0.19*).

With respect to search hits clicked, users performed the activity much more frequently on WIGI (432 times) than they did on the browser (179 times) as shown in Table 17. The difference between the number of search hits clicked on WIGI and the browser was significant according to ANOVA (*F (1, 58) = 34.37, p<0.0001*). The activities performed for finding information show a significant difference only in the case of clicking search hits. Users behaved similarly in the cases of clicking links on web pages and submitting queries on websites. They also submitted almost the same number of queries to search engines in the cases of using both WIGI and the browser. Finally, the data show that users of both WIGI and the browser rarely used the find-on-page feature. There was no significant difference between the data recorded on WIGI and the data recorded on the browser for the use of the find-on-page search feature.

6.8.5.Other Activities

In addition to the activities related directly to re-finding information, handling multiple sessions, and managing and organizing information, other activities were logged during the study as shown in Table 18.

Table 18. Other activities.

Participant	Bookmark created	Bookmark opened	File created	File opened	Tab closed	Thumbnail Removed (WIGI)
	Browser					
P1	4	1	0	1	0	0
P2	0	0	2	1	1	1
P3	0	0	1	0	3	0
P4	0	0	2	1	3	3
P5	0	0	1	0	0	2
P6	0	0	1	0	0	3
P7	0	0	1	1	4	9
P8	0	0	0	0	0	8
P9	0	0	2	2	2	3
P10	0	0	1	0	1	3
P11	0	0	1	1	3	5
P12	0	0	2	0	0	10
P13	0	0	0	1	7	4
P14	0	0	1	1	0	1
P15	0	0	1	1	0	6
P16	2	2	0	0	0	2
P17	0	0	1	1	0	2
P18	6	6	1	0	0	9
P19	0	0	1	1	0	1
P20	0	0	1	1	0	1
P21	0	0	2	0	0	1
P22	0	0	1	1	0	4
P23	0	0	1	0	0	7
P24	0	0	1	1	1	3
P25	0	0	1	1	0	3
P26	0	0	1	1	0	10
P27	0	0	1	1	0	10
P28	0	0	1	1	0	2
P29	0	0	2	5	0	6
P30	3	1	2	0	0	3
Total	**15**	**10**	**33**	**24**	**25**	**122**
Mean	**0.5**	**0.33**	**1.1**	**0.8**	**0.83**	**4.07**
SD	**1.41**	**1.15**	**0.61**	**0.96**	**1.64**	**3.17**

All users of WIGI indicated that thumbnails associated with titles of web pages (linked to URLs) worked better than the use of tabs. Twenty six users (26/ 30) indicated that they did not use browser tabs with WIGI because they considered the thumbnail view in the reference tracking area equivalent to the use of tabs, yet with additional benefits. Compared to tabs, thumbnails were always available on the display and could be clicked to see the page content without the loss of viewing other available thumbnails representing other pages which happens in the case of using tabs. On the browser, closing a tab means that the user, for some reason, did not need to have the page open. On WIGI, this is equivalent to removing a thumbnail from the reference tracking area.

The use of thumbnails was shown to be more frequent than the use of tabs. On WIGI, users opened 430 thumbnails (7.98% of all activities). They opened 235 tabs (4.30% of all activities) while using the browser. Users of WIGI removed an average of four thumbnails denoting a high level of activity compared to only 0.8 tabs closed on the browser. The use of thumbnails was more frequent in the case of WIGI than the use of tabs with the browser. Users used thumbnails as an alternative to tabs and never used multiple tabs on WIGI.

6.8.6.A Comparison of Actions per Activity (WIGI vs. Browser)

The number of steps (actions) required to perform an activity on WIGI was compared to those taken to complete the same activity while using the browser and substantial differences were observed. As shown in Table 19, users needed only one click to re-find information using the thumbnail view in the reference tracking area compared to multiple steps on the browser even when the page was already open. To handle multiple sessions using the *save gathered/retrieve task* features on WIGI, the user needed no more than two clicks for saving and two for restarting the task. On the other hand, it is not possible to predict or determine how many clicks were needed to keep all the task information and restart the task at later sessions while using the browser especially when other applications or tools were used. This is because the user may have different strategies to handle multiple sessions in the case of the browser. For managing and organizing information on WIGI, the user needed one click to copy and paste (i.e. drag the information), format, or type the information. On the browser, however, the user needed more clicks to perform the same activities.

Table 19. Steps required for activities on WIGI vs. browser.

Subtask	WIGI			Browser		
	Activity		Actions	Activity	Actions	
Re-finding	Revisit link		Find link in reference tracking area	Revisit link	Copy link	
					Switch application	
			Click link		Past link	
					Click address bar	
Handling Multiple Sessions	Keep task		Hit save gathered	Keep dispersed parts of the task	One or more of:	Save page
			Enter task name (in the case of a new task)			Save session of tabs
						Save file
						Email
	Retrieve Task		Select task	Retrieve scattered parts of the task	One or more of:	Open saved page
						Open saved session of tabs
						Search for file
						Open file
						Open email
Managing and Organizing	Copy and past		Drag and drop	Copy and past	Copy information	
					Paste information	
	Format		Apply formatting	Format	Apply formatting	
	Type		Type information	Type	Type information	
	Create a reference		Embed reference	Create a reference	Copy link	
					Paste like	
	Search	Web search	Submit query	Web search	Submit query	
		On page search	Browser find	On-page search	Browser find	
		Site search	Submit query	Site search	Submit query	

6.8.7.Post-Task Questionnaire Data

Even though the main focus of the study was using the logged data to investigate the effectiveness of specific features embedded in WIGI, users' opinions about the task and the tools used were recorded in the post-task questionnaires. There were two questionnaires used, one for WIGI and another one for the browser. The only difference was that the WIGI version of the questionnaire asked an additional question regarding the features embedded in WIGI. The questionnaires are shown in Appendices E and F.

On WIGI, all users rated their completion of the task on a 7-point Likert scale. Of the participants, 76.7% (23/30) chose the highest score, seven, with respect to satisfying the completion of the task while 23.3% (7/30) chose six on the scale. Users were asked to rate their confidence in the requirements they gathered for the task. Of the participants, 46.7% (14/30) chose the highest level, seven, on the scale; 50% (15/30) chose six; and the remaining one user chose five. Moreover, users were asked about their evaluation of particular features in WIGI as shown in Table 20.

Table 20. User ratings of the effectiveness of particular features in WIGI.

Fearure	**Very effective**						**Not effective at all**
Ability to keep all the task information as one unit.	80.0% (24)	16.7% (5)	3.3% (1)	0.0% (0)	0.0% (0)	0.0% (0)	0.0% (0)
Ability to retrieve all the task information as one unit.	72.4% (21)	24.1% (7)	3.4% (1)	0.0% (0)	0.0% (0)	0.0% (0)	0.0% (0)
Thumbnails for tracking references.	50.0% (9)	50.0% (15)	0.0% (0)	0.0% (0)	0.0% (0)	0.0% (0)	0.0% (0)
Ability to edit and format information along with browsing and searching.	70.0% (21)	26.7% (8)	3.3% (1)	0.0% (0)	0.0% (0)	0.0% (0)	0.0% (0)
Ability to embed references while editing information.	83.3% (25)	13.3% (4)	3.3% (1)	0.0% (0)	0.0% (0)	0.0% (0)	0.0% (0)
The use of WIGI as a whole.	46.7% (14)	36.7% (11)	13.3% (4)	3.3% (1)	0.0% (0)	0.0% (0)	0.0% (0)

Users rated the use of thumbnails for tracking references in the reference tracking area as very effective using the highest two levels of the scale as shown in Table 20. The ability to edit and format information along with browsing and searching was also rated as very effective by 70.0% of the users (21/30). Embedding references achieved the highest rating of seven on the scale given by 83.3% (25/30) of the users. The ability to keep the task information integrated achieved the highest rate of seven from 80.0% (24/30) of the users. Re-finding the task information as one unit was rated with the highest level on the scale by 72.4% (21/30) of the users. The use of WIGI as a whole was given the highest rate by 46.7% (14/330) of the users followed by 36.7 (11/30) users giving the second highest level on the scale. Only very few individuals selected the third level on the scale to rate those features in WIGI.

Users provided comments regarding future improvements that they would like to see in WIGI. Those comments were classified by the researcher as shown in Table 21. The table shows the constructive feedback comments left by the participants as well as the positive comments participants provided after finishing the tasks. The constructive feedback included mostly individual comments and they were of personal preferences. The positive comments included that WIGI was effective and useful and that users needed no other tools to finish the task. The total number of comments provided was 23 comments. Seven users either did not leave any comments or left comments that were too general or unrelated. Therefore, those comments were omitted from Table 21.

Table 21. User Comments (WIGI).

Constructive Feedback	
Comment topic	**Number of comments**
Summary along with thumbnails	1
Open link directly in browser without having to click the special link	2
Keep track of all pages opened not just the ones clicked in search results	1
More browsing area	2
Other (non-relevant comments)	3
Total	**9**
Positive Comments	
Comment topic	**Number of comments**
No switching was helpful with typing	2
Very effective and useful tool over the browser	7
Everything was great. I needed nothing	4
Embedded editor was a great idea	1
Total	**14**

On the browser, users had the same post-task questionnaires excluding the question that asked about the perceived effectiveness of certain features in WIGI (Question 5). Regarding completing the tasks, 70% of the users (21/30) answered with the highest score, seven, on the Likert scale used. Four users (13.3%) chose the score of six while five users (16.6%) chose five on the scale. With respect to the user confidence in completing the task requirements, only 13% of the users (4/30) chose the highest score, seven, on the scale. In addition, 40% of the participants (12/30) chose the following score of six while another 40% (12/30) chose the score of five on the scale. The remaining two users (7%) chose the score of four on the scale.

Twenty six participants (86.7%) indicated that they used text files to save the task information while the remaining four users (13.3%) indicated using emails. The results of this section in the post-task questionnaire agree with the actual data logged in the study. To manage the task information, 28 participants used text editors while the remaining two used emails drafts. Four users used bookmarks to keep references to task information while two users saved complete pages. Consequently, different strategies were used to complete the task while managing, keeping, and re-finding information.

Table 22. Comments in the browser post-study questionnaires.

Comment Topic	**Number of comments**
Re-finding all information together.	1
Organization of my previous session information.	1
More effective copying and pasting.	1
Easier way to bookmark.	1
Summaries of the page I viewed.	1
Keep notes on browser.	1
Better organization of task information from previous session. Could not find file.	1
Keeping the main information in useful pages.	1
Better way to organize tabs.	1
Task problem.	1
Ability to see search history.	2
Tab saving feature.	2
No problems encountered.	4
Not having to switch by embedding an editor into the browser.	5
Using WIGI.	7
Total	**30**

While using the browser, users provided different comments regarding issues they had with the tools they used and features they wanted to use that were not available during the study. The comments were categorized as shown in Table 22. Most of the comments concerned individual issues. The most frequent comment (7/30) indicated that the user wanted to use WIGI as an alternative to the browser and other tools used. Embedding an editor into the browser was also stated five times (5/30). Four users (4/30) indicated that they had no problems with the tools they used for the task while on the browser. Users who indicated using WIGI as an alternative are of those who already finished the task on WIGI before completing the post-task questionnaire with the browser.

By comparing the case of WIGI to the browser with respect to the user confidence in completing the task, there was a significant difference between the number of users (14/30) who were confident in satisfying the task requirements while using WIGI and those (4/30) who were confident in satisfying the task requirements while using the browser (*z-test, z=3.02, p<0.003*). Users of WIGI provided 23 comments, of which 14 comments were positive. The six remaining comments concerned issues users had with WIGI as shown in Table 21. While using the browser, users provided 26 negative comments (26/30), of which seven comments (7/26) expressed preference to use WIGI. The difference between the number of issues on which users commented while using WIGI and the number of issues they had with the browser according to their comments was significant (*z-test, z = - 3.99, p<0.0001*).

6.9. Study Limitations

Similar to the previous study discussed in Chapter 5, all users in this study were computer science students who did not reflect all information gatherers on the web. In addition, the study used simulated tasks that may not reflect all possible realistic information gathering tasks.

6.10. Discussion

The previous study provided three main recommendations concerning: re-finding information for the task, handling multiple sessions, and managing and organizing information. Based on those recommendations, the features illustrated in Section 6.3 were implemented in a prototype and a user study was conducted to compare those features to the current state of information gathering on the web represented by the browser and other complementary applications and tools.

In this study, users indicated concerns with issues that confirmed the recommendations developed in the previous study. However, the data logged in the current study showed that the priority of consideration given to the subtasks as implied in the user responses do not reflect the frequency of activities performed for each subtask. For example, managing and organizing information had fewer issues from the user's perspective while the data showed that this subtask had more activities logged than the subtasks of re-finding information or handling multiple sessions. For answering the second question, different activities were logged and the user responses were analyzed with respect to the three issues indicated in the question.

Re-finding information while gathering information on the web was investigated by implementing the ability to keep track of the references opened by the user (i.e. web pages). Keeping track of references was done by keeping the title of the page opened from the search hits associated with the link and thumbnail of the page in the reference tracking area. The results showed that users performed significantly more re-visiting activities for re-finding information on WIGI than they did on the browser. Re-finding on the browser was done through copying and pasting links kept in text files into the address bar of the browser.

In the case of WIGI users indicated that it was more convenient to see the references in the context of the information being gathered. They also indicated that clicking pages kept as thumbnails in the reference tracking area to reopen them was very effective. On the browser, when users asked, they indicated that copying links from the text files where the information was gathered and re-opening the pages was demanding. Bookmarks were rarely used by users as shown above. Users indicated that using bookmarks would take them out of the context of the task since bookmarks were kept separate from the rest of the information being gathered. Only 10 of the 15 bookmarks created were re-opened during the second session of the task.

To handle multiple sessions while using WIGI, the *save gathered* feature allowed users to keep all the information of the task including all objects collected (text, images, links… etc.) along with the references tracked under one name in a searchable list. The task was saved under one label so that the user could reopen the task and go back to exactly the same context of the task where they left in the last session. This feature was used by all users of WIGI in the study and none of the users used any other form of task keeping such as files.

On the browser, the use of text files to keep information over multiple sessions was more frequent than the use of other methods such as bookmarking and using emails. The ***z*** test results showed that the proportion of cases using text files compared to the proportion of cases using bookmarks was significant (*z = 5.68, p<0.0001*). The same results was achieved when comparing the use of text files to the use of emails (*z = 5.68, p<0.0001*). All users of WIGI used the *save gathered* feature to keep the task information and the *retrieve task* feature to go back to the task context in subsequent sessions.

Users (24/30) indicated that they found it very effective to keep the task information integrated for working on the same task over multiple sessions as implemented in WIGI. They found retrieving the task information along with references helpful in handling multiple sessions since it preserved the context of the task. On the browser, five users (5/30) lost the information they kept in the first session while trying to complete the second part of the task in the second session even though the time gap between the two sessions was only one day. They indicated that keeping the task files in the file system hierarchy had always caused the loss of information over time. Users of bookmarks did not make use of one third of the bookmarks they created. They indicated that they forgot about the bookmarks they had created during the first part of the task.

With respect to managing and organizing information, WIGI activities were different. The main features which made the difference were: copying and pasting information, typing information, formatting information, and embedding references.

Users performed copying and pasting more frequently on WIGI than they did while using the browser. The difference was significant as discussed above. Users indicated that it was easy to copy and paste information on one display in the case of WIGI while this activity required switching between the browser and other tools used to gather the task information in the case of the browser. Copying and pasting from web pages on WIGI required the user to drag and drop objects from the browsing area or the search result hits into the editor. This feature may have motivated users to perform this activity more frequently on WIGI. On the browser, however, users had to copy the information from open pages on the browser or from result hit summaries and then paste the information into the text editor or email draft being edited. The copying and pasting activity required the use of two different application and further formatting in some cases. It may have also lead to more typing in the case of the browser perhaps to avoid the additional step of switching between applications to complete pasting the information.

Typing information was one of the organizing and managing activities that showed a difference between WIGI with its features and the browser with other complementary applications used for the given tasks. This activity was performed significantly more frequently using the browser than it was performed while using WIGI. It was shown in the data that the difference with regard to the frequency of typing on WIGI and the browser cannot be attributed to the individual differences. It may be that the tool used is what caused the difference. By comparing typing to copying and pasting, it seems that users typed more in the case of the browser because they were less motivated to copy and paste information.

Users performed significantly more formatting activities using WIGI than they did in the case of using the browser (521 vs. 51 activities). Users of WIGI used the embedded editor which kept the original formatting of any objects that were copied and pasted (or dragged) from web pages. Though, users were motivated to perform more formatting activities through the use of the formatting features provided with the editor. They indicated that the layout of the editor in the middle of the display and the ability to quickly perform any desired formatting activity using the features provided made it easy to format the objects collected.

Even though 50% of the users (15/30) used MSWord while using the browser to complete the tasks, they rarely formatted the information they collected. The reason given by participants was the need to switch between the browser and the text editor. The current state of the web gathering process did not motivate users to format the information. Users who used Notepad and email drafts did not do any formatting to the information. The results indicate that WIGI, although the reasons remain unclear, did motivate users to format the information more frequently. In addition to what the users indicated, it seems that combining editing with browsing and searching may have allowed users to see (on one display) the information being gathered and the sources of the information with the original formatting. Therefore, they may have been indirectly motivated to perform formatting significantly more frequently on WIGI.

A part of managing and keeping information was embedding references into the information gathered. With respect to the number of references embedded, there was no difference between WIGI and the browser. Users embedded references in both cases with similar frequency to keep track of sources of information they gathered. In the case of WIGI, references were embedded using a feature provided in the control bar that adds the reference into the editor where the cursor is located. In the case of the browser, the user had to use text editors to copy the URL of the reference into the text file or email. Some references were kept as bookmarks separate from the information being gathered.

Embedding references on WIGI was done either involving thumbnails of the pages being referenced or excluding thumbnails. The difference between users who excluded the thumbnails and those who kept them was not significant. In both case, it was easier and less demanding to embed references using WIGI than it was using the browser where users had to use two applications most of the time. Of the 30 participants, 25 users (83.3%) indicated that it was very effective to embed references using WIGI. Even though users embedded almost the same number of links in the case of using the browser, they did so because it was the only way to keep track of those references even with: the use of two applications, the copying and pasting, and the switching that was required.

Organizing and managing information on WIGI was different with respect to formatting the task information, copying and pasting the information, and typing the task information. The difference was significant in those cases and WIGI changed the way users managed and organized the task information. Users did more formatting on WIGI and performed more copying and pasting activities. They typed information on fewer occasions though. The results indicate that users found WIGI to be helpful in gathering the information and providing the required formats.

To summarize, the results of the study showed that:

1. Users, although indicated concerns with issues similar to those connected to the recommendations from the earlier study, their focus was more on the subtask of re-finding information and the process of handling multiple sessions in the

comments they reported. Based on the activities recorded, the last study showed that managing and organizing the information for the task was more important than other subtasks and should be considered for further investigations.

2. According to the number of activities recorded for each subtask, managing and organizing the task information is still a very important task, and it requires further investigations.
3. The features investigated in the study demonstrated differences in the way users performed the tasks as follows:
    a. Re-finding information was more effective on the prototype interface compared to the other strategies users of the browser used.
        i. Reference tracking through the use of thumbnails (associated with URLs and titles of pages visited) was used significantly more than keeping track of links using different strategies while using the browser (mainly copying and pasting links into text files and emails).
        ii. All users rated the reference tracking feature embedded in WIGI as effective.
    b. Handling multiple sessions was more effective on WIGI compared to the case of the browser.
        i. The *save gathered* and *retrieve task* features prevented the loss of information over multiple sessions. Of the participants, 20% (5 participants) lost the task information from the previous session while using the browser which never occurred when using WIGI.
        ii. These *save gathered* and *retrieve task* features were the only strategies users needed to handle multiple sessions on WIGI.
        iii. Users made use of bookmarks for keeping and never came back to open 33% of the bookmarks they created on the browser.
        iv. The difference between users of WIGI who used the *save gathered* and *retrieve task* features and those who used text files (the most frequent strategy used for handling multiple sessions on the browser) was statistically significant.

c. Managing and organizing information was more effective using the prototype WIGI than using the browser and other complementary applications.

   i. Having the embedded editor along with the search and browsing areas on one display lead to:

      1. Significantly more copying and pasting of information on WIGI than on the browser. Copying and pasting are core activities for collecting and managing the task information
      2. Significantly more formatting activities on WIGI to manage and organize the information for the task than in the case of using the browser.
      3. Significantly fewer typing activities on WIGI than in the case of using the browser.

   ii. Twenty nine users (96.7%) rated the ability to edit, format, search, and browse the information on one display as effective. They indicated that this feature eliminated the need for switching among different applications.

   iii. Twenty five participants (83.4%) rated the ability to embed references into the editor as effective.

Table 23 provides a comparison showing the research questions and how those question were answered in the study. As shown in Table 23, except for embedding references as a part of the subtask of information management and organization, the remaining features achieved a substantial level of success over the current state of information gathering on the web. Re-finding by revisiting links from previous sessions for the task was enhanced by keeping track of search hits clicked. Handling multiple sessions was improved by eliminating the loss of information transferred over to subsequent sessions. Managing and organizing information was improved by eliminating switching among tools and application. In addition, the user performed fewer typing activities and more pasting activities by providing the editor on the browser along with search capabilities.

Table 23. Study results (summarized).

Research Question		**Answers**
Do the behavioural characteristics of users performing information gathering tasks on the web confirm the recommendations developed in the earlier study? This question is answered by investigating: a. The data logged during the study for each subtask examined. b. The user responses to issues regarding their behaviour while performing information gathering tasks.		- The study had more activities concerning the subtask of managing and organizing information than re-finding or handling multiple sessions. - The user responses to issues regarding subtasks that require more investigations agreed with the recommendations from the last study yet with more focus of the user on the subtask of re-finding information and handling multiple sessions in the comments reported. - The previous study indicated managing and organizing the task information as the most important subtask for investigations based on the number of activities performed in this subtask. The same conclusions apply in this study based on the number of activities recorded.
What is the effectiveness of each of the following specific features on subtasks completed by users as parts of the information gathering task on the web?	Keeping track of references for web pages using thumbnails accumulated altogether in a reference tracking area compared to conventional methods such as copying and pasting links into text files in the case of the subtask of re-finding information.	- Reference tracking through the use of thumbnails (associated with URLs and titles of pages visited) was used more significantly than keeping track of links using different strategies while using the browser. - All users rated the reference tracking feature embedded in WIGI as effective.
	Keeping and retrieving the task information integrated compared to keeping and later retrieving parts of the task using conventional strategies such as saving pages, saving information in files, bookmarking, and so forth in the case of the subtask of handling multiple sessions.	- The *save gathered* and *retrieve task* features prevented the loss of information over multiple sessions. - These two features were the only strategies users needed to handle multiple sessions. - Users made use of bookmarks for keeping and never came back to open 33% of the bookmarks they created while using the browser. - The difference between users of WIGI who used the *save gathered* and *retrieve task* features and those who used text files to handle multiple sessions on the browser was statistically significant.
	Using one application for searching, browsing, and editing compared to the use of multiple applications (i.e. the web browser and a text editor) in the case of the subtask of managing and organizing information.	- Having the embedded editor along with the search and browsing areas on one display lead to: * Significantly more copying and pasting of information on WIGI than on the browser. * Significantly fewer typing activities on WIGI than in the case of using the browser. * Significantly more formatting activities on WIGI to manage and organize the information for the task than in the case of using the browser. * Embedding references on WIGI, although rated as very effective, was not used significantly more than copying and pasting references in the case of the browser. - Twenty nine users (96.7%) rated the ability edit, format, search, and browse the information on one display as effective. They indicated that this feature eliminated the need for switching among different applications. - Twenty five participants (83.4%) rated the ability to embed references into the editor as effective.

6.11. Summary

This chapter discussed a study conducted to investigate particular features designed and implemented based on the recommendations of the earlier study. The study investigated the effectiveness of those features on how users complete information gathering tasks on the web. The features concerned three subtasks: re-finding information, handling multiple sessions, and managing and organizing the task information. The study showed that those features achieved significant improvements over the ordinary web browser and other tools used to complete the task.

CHAPTER 7 FINAL DISCUSSION AND CONCLUSIONS

The research went through different stages that started with modeling the task of information gathering by identifying its underlying subtasks. The model created for the task of information gathering depicted the task accomplishment process and the relationships among the subtasks identified. Features of the task of information gathering were associated with each applicable subtask to help with further investigations. The model was initially created based on activities of the information gathering task that were investigated in previous research.

After conducting two studies that examined two subtasks of the initial model—finding information sources and finding information—the model was further adjusted. The subtasks of finding information and finding further related information were merged into one subtask. The subtask of keeping information was considered as part of the subtask of managing and organizing information. Finally, after conducting the exploratory study discussed in Chapter 5, the model had its final refinement by adding the subtask of handling multiple sessions which was considered as a separate subtask based on the activities performed in the study and the issues revealed to be of concern to the user.

The changes applied to the model helped with creating a final framework of the subtasks that comprise the task of information gathering. The framework helped with developing a definition of the task of information gathering that went beyond a simple description of the task. It also helped with identifying the core subtasks in information gathering, revealing tools and strategies users use to accomplish the subtasks investigated, uncovering difficulties users encounter with each subtask, and recommending which subtasks required further investigations. The model had other subtasks that may be investigated in future research. Moreover, the model may help with studying other kinds of tasks on the web using the same approach.

Based on the model developed and the features identified, the task of information gathering is defined as a combination of multiple subtasks each of which consists of different activities. Information gathering is complex and highly search-reliant. It requires collecting different kinds of data possibly from different sources. It may necessitate the use of multiple applications, and it may require multiple sessions to complete. The definition of the task helped with the design and implementation of different features that were investigated in this research.

In order to understand the concept of information gathering and further investigate the task, two preliminary studies were conducted as part of this research. The first study investigated the subtasks of finding information and information sources on the web. The second study investigated how users locate further related information (continuing the subtask of finding information) by following the hyperlink connectivity on the web graph. These two studies helped with: refining the information gathering task model that was initially created based on the literature review, constructing simulated tasks for further studies, and developing a definition of the task of information gathering.

After creating the task model and identifying the core subtasks that comprise the information gathering task, the third study investigated the subtask of re-finding information in addition to the subtask of managing and organizing information. The choice of these two subtasks was recommended as a result of the two preliminary studies in addition to a questionnaire-based study discussed in Alhenshiri, et al. (2011). The investigation recommended building support for three subtasks involved in information gathering: re-finding information, handling multiple sessions, and managing and organizing information. The information gathering task model was further refined to involve the subtask of handling multiple sessions.

The last investigation concerned evaluating features embedded in the web browser to support the subtasks of: re-finding information, handling multiple sessions, and managing and organizing information. The features intended for evaluation were investigated against the use of features available in the ordinary browser. Other complementary applications needed by users gathering information on the web were available for use during the study. The results of the investigations showed that current web tools—mainly the web browser—are not sufficient for the task of information gathering having the features that are already implemented in those tools. The features added and investigated in this research achieved significant improvements over the use of the current browsing model.

Future work may consider implementing the features investigated in this research. Research may further study other subtasks in the information gathering task. For instance, since information gathering involves comparing information from different sources, investigating the subtask of comparing information, reasoning, and decision making may yield findings that help with improving how user complete this type of task. Those findings may consider different designs of the current browsing model. There are other subtasks and characteristics of the task of information gathering that require further investigations such the subtask of managing and organizing information. Research should continue investigating the activities underlying each subtask involved in information gathering on the web due to the ongoing changes in the user behaviour that result from the continuous changes to the web.

In addition to the benefits of this research in supporting information gathering tasks for users working on desktop computers, the research may continue with small-screen devices. The use of smart phones and tablets has become dominant in almost every aspect of life. In the first quarter of 2012 alone, Apple[13] shipped 15 million iPad devices. By the year 2015, there will be 7.4 billion 802.11n devices in the market according to Forbes[14]. According to Vertic[15], enterprise tablet adoption will grow by almost 50% per year, and by 2015, mobile application development projects will outnumber native personal computer projects by a ratio of 4-to-1.

Different interfaces have been designed to work on small-screen devices. However, most of the web applications utilized on tablets and smart phones are those used on desktop computers. Accomplishing complete tasks rather than individual activities on small-screen devices require further investigations with regard to aspects in the design of user interfaces on those devices. The problems associated with completing a user task on the web—particularly information gathering—has not yet been fully investigated even on devices with larger displays such as desktop computers. Further investigations of user tasks on the web in the case of small-screen devices may benefit larger numbers of users.

[13] http://www.apple.com
[14] http://www.forbes.com
[15] http://www.vertic.com

BIBLIOGRAPHY

Adar, E., J. Teevan, and S. Dumais. "Large Scale Analysis of Web Revisitation Patters." *ACM Conference on Human Factors in Computing Systems.* Florence, Italy, 2008. 1197-1206.

Alhenshiri, A, C Watters, M Shepherd, and J Duffy. "Information Gathering within Websites: Visualized Links for Navigation (VLN)." *ISDA2010.* Cairo, Egypt: IEEE, 2010a. 766-771.

Alhenshiri, A., and J. Duffy. "User Studies in Web Information Retrieval: User-Centered Measures in Web IR Evaluation." *2010 IEEE 24th International Conference on Advanced Information Networking and Applications .* Perth, Australia: IEEE, 2010. 632-637.

Alhenshiri, A., and W. J. Blustein. "Exploring Visualization in Web Information Retrieval." *International Journal for Internet Technology and Secured Transactions (IJITST)*, 2010a.

Alhenshiri, A., and W. J. Blustein. "Utilizing Visualization for Improving Web Search Effectiveness." *the International Conference on Information Society (i-society 2010).* London, UK: IEEE, 2010b.

Alhenshiri, A., C. Watters, and M. Shepherd. "Exploring the Concepts of Visualization, Clustering, and Re-finding in Web Information Gathering Tasks: a Survey." *2nd IEEE Symposium on Web Society.* Beijing, China: IEEE, 2010c.

Alhenshiri, A., C. Watters, and M. Shepherd. "Improving Web Search for Information Gathering: Visualization in Effect." *4th Workshop on Human-Computer Interaction and Information Retrieval (HCIR2010).* New Brunswick, NJ, USA: Microsoft, 2010d.

Alhenshiri, A., C. Watters, and M. Shpherd. "VISTO: for Web Information Gathering and Organization." *4th Workshop on Human-Computer Interaction and Information Retrieval, IIiX2010 Symposium.* New Brunswick, NS, USA: Microsoft, 2010e.

Alhenshiri, A., M. Shepherd, and C. Watters. "Investigating User Behaviour during Web Search as Part of Information Gathering." *Hawaii International Conference on System Sciences (HICSS44).* Koloa, Kauai, Hawaii, USA: IEEE, 2011.

Alhenshiri, A., M. Shepherd, C. Watters, and J. Duffy. "Web Information Gathering Tasks: a Framework and Research Agenda." *the International Conference on Knowledge Discovery and Information Retrieval (KDIR2010).* Valencia, Spain: Springer, 2010f.

Alhenshiri, A., S. Brooks, C. Watters, and M. Shepherd. "Augmenting the visual presentation of web search results." *The 5th International Conference on Digital Information Management.* Thunder Bay, ON, Canada, 2010. 101-107.

Alhenshiri, A., Shepherd, M., Watters, C. "Effective Information Gathering on the Web." *the iConference.* Toronto, ON, Canada, 2012b. 415-416.

Alhenshiri, A., Shepherd, M., Watters, C., Duffy, J. "User Search Behaviour During Web Information Gathering Tasks." *the 8th International Conference on Web Information Systems and Technologies.* Porto, Portugal, 2012d. 323-331.

Alhenshiri, A., Watters, C., Shepherd, M., Duffy, J. "Building Support for Web Information Gathering Tasks." *the 45th Hawaii International Conference on System Sciences.* Grand Wailea, Maui, Hawaii, 2012a. 1687-1696.

Alhenshiri, A., Watters, C., Shepherd, M., Duffy, J. "Investigating Web Information Gathering Tasks." *the ASIS&T75 annual meeting.* Baltimore,MD, 2012c. 26-30.

Alonso, O., and R. Baesa-Yates. "Alternative Implementation Techniques for Web Text Visualization." *1st Latin American Web Congress.* California, USA: ACM Press, 2003. 202-204.

Amin, A. "Establishing Requirements for Information Gathering Tasks." *TCDL Bulletin of IEEE Technical Committee on Digital Libraries* 5, no. 2 (2009): 1-10.

Au, P., M. Carey, S. Sewraz, Y. Guo, and S. Rugers. "New Paradigms in Information Visualization." *23rd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval.* Athens, Greece, 2000. 307-309.

Badesh, H., and W. J. Blustein. "VDMs for Finding and Re-finding Web Search Results." *2012 iConference.* Toronto, ON, Canada: ACM, 2012. 419-420.

Bagchi, A., and G. Lahoti. "Relating Web Pages to Enable Information-Gathering Tasks." *ACM Conference on Hypertext and Hypermedia.* Torino, Italy, 2009. 100-118.

Bederson, B. B., J. D. Hollan, J. Stewart, D. Rogers, A. Druin, and D. Vick. "A Zooming Web Browser. SPIE Multimedia Computing and Networking." *SPIE Multimedia Computing and Networking.* 1996. 260-271.

Bell, D. J., and I. Ruthavan. "Assessments of Task Complexity for Web Searching." *26th European Conference on Information Retrieval.* Sunderland, UK, 2004. 57-71.

Berkun, S. "The Explorer Bar: Unifying and Improving Web Navigation." *IFIP TC.13 International Conference on Human-Computer Interaction (Interact'99).* Amsterdam, the Netherlands, 1999. 156-162.

Bonnel, N., A. Cotarmanac'h, and A. Morin. "Meaning Metaphor for Visualizing Search Results." *9th International Conference on Information Visualization.* London, UK, 2005. 467-472.

Bonnel, N., V. Lemaire, A. Cotarmanac'h, and A. Morin. "Effective Organization and Visualization of Web Search Results." *24th IASTED Internation Multi-Conference on Internet and Multimedia Systems and Applications.* Innsbruck, Austria, 2006. 209-216.

Broder, A. "A Taxonomy of Web Search." *ACM SIGIR Forum* 36, no. 2 (2002): 2-10.

Büttcher, S., Clarke, C. L., Cormack, G. V. *Information Retrieval, implementing and evaluating search engines.* MIT Press. , 2010.

Bystrom, K., and P. Hansen. "Conceptual Framework for Tasks in Information Studies." *Journal of the American Society of Information Science and Technology* 56, no. 10 (2005): 1050-1061.

Cai, D., X. He, Z. Li, W. Ma, and J. Wen. "Hierarchical Clustering of WWW Image Search Results Using Visual, Textual, and Link Information." *12th Annual ACM International Conference on Multimedia.* New York, NY, USA, 2004. 951-959.

Capra, R. G., and M. A. Perez-Quinones. "Factors and Evaluation of Re-finding Behaviour." *SIGIR 2006 Workshop on Personal Information Management.* Seattle, Washington, USA: ACM, 2006. 10-11.

Card, S. K., J. D. Mackinlay, and B. Shneiderman. *Readings in Information Visualization: Using Vision to Think.* San Francisco, CA, USA: Morgan Kaufman Publishers, 1999.

Carpineto, A., S. Osinski, G. Ramano, and D. Weiss. "A Survey of Web CLustering Engines." *Computing Surveys* 41, no. 3 (2009): Article No. 17.

Choo, C. W., B. Detlor, and D. Turnbull. "A Behavioral Model of Information Seeking on the Web--Preliminary Results of a Study of How Managers and IT Specialists Use the Web." *The Annual Meeting of the American Society for Information Science.* Pittsburgh, PA, USA: ASIS, 1998. 25-29.

Choo, C. W., B. Detlor, and D. Turnbull.. "Working the Web: An Empirical Model of Web Use ." *33rd Hawaii International Conference on System Sciences.* Maui, Hawaii, 2000. 7064.

Cockburn, A., B. McKenzie, and M. JasonSmith. "Evaluating a Temporal Behaviour for Web Browsers' Back and Forward Buttons." *11th International WWW Conference.* Honolulu, Hawaii, USA, 2002. 1.

Cutrell, E., S. T. Dumais, and J. Teevan. "Searching to Eliminate Personal Information Management." *Communications of the ACM* 49, no. 1 (2006): 58-64.

Dearman, D., M. Kellar, and K. N. Truong. "An Examination of Daily Information Needs and Sharing Opportunities." *2008 ACM Conference on Computer Supported Cooperative Work.* San Diego, CA, USA, 2008. 679-688.

Di Giacomo , E., W. Didimo, L. Grilli, G. Liotta, and P. Palladino. "WhatsOnWeb+: An Enhanced Visual Search Clustering Engine." *Visualization Symposium (PacificVIS '08).* Perugia, Italy, 2008. 167-174.

Doll, W. J., and G. Torkzadeh. "The Measurement of End User Computing Satisfaction." *MIS Quarterly* (Management Information Systems Research Center, University of Minnesota) 12, no. 2 (1988): 259-274.

Dong, L., C. Watters, M. Shepherd, and J. Duffy. "An Examination of Genre Attributes for Web Page Classification." *41st Annual Hawaii International Conference on System Sciences.* Waikoloa, Hawaii, USA, 2008. 133.

Dörk, M., C. Williamson, and S. Carpendale. "Towards Visual Web Search: Interactive Query Formulation and Search Result Visualization." *WWW Workshop on Web Search Result Summarization and Presentation.* Madrid, Spain, 2009. 100-104.

Efron, M., J. Elsas, G. Marchionini, and J. Zhang. "Machine Learning for Information Architecture in a Large Government Website." *4th ACM/IEEE-CS Joint Conference on Digital Libraries.* Tuscon, AZ, USA, 2004. 151-159.

Ellis, D. "A Behavioural Approach to Information Retrieval System Design." *Journal of Documentation* 45, no. 3 (1989): 171-212.

Ellis, D., and M. Haugan. "Modelling the Information Seeking Patterns of Engineers and Research Scientists in an Industrial Environment." *Journal of Documentation* 53, no. 4 (1997): 384-403.

Ellis, D., D. Cox, and K. Hall. "A Comparison of the Information Seeking Patterns of Researchers in the Physical and Social Sciences." *Journal of Documentation* 49, no. 4 (1993): 356-369.

Elseweiler, D., and I. Ruthavan. "Towards Task-based Information Management Evaluation." *ACM SIGIR Conference on Research and Development in Information Retrieval.* Amsterdam, The Netherlands: ACM, 2007. 23-30.

Ferizis, G., and P. Bailey. "Towards Practical Genre Classification of Web Documents." *15th International Conference on World Wide Web.* Edinburgh, Scotland, 2006. 1013-1014.

Ferragina, P., and A. Gulli. "A Personalized Search Engine Based on Web-snippet Hierarchical Clustering." *14th International Conference on World Wide Web.* Chiba, Japan, 2005. 801-810.

Fitt, P. M. "The Information Capacity of the Human Motor System in Controlling the Amplitude of Movement." *Journal of Experimental Psychology* 47, no. 6 (1054): 381-391.

Friendly, M. *Milestones in the History of Thematic Cartography, Statistical, Graphics, and Data Visualization.* n.d. http://www.math.yorku.ca/SCS/Gallery/milestone/ (accessed March 2009).

Grayson, T. D., R. A. Grayson, and G. E. Hedrick. "A web information organization and management system (WIOMS)." *16th ACM Symposium on Applied Computing (SAC2001).* Las Vegas, USA: ACM, 2001. 565-574.

Grewal , R. S., M. Jackson, P. Burden, and J. Wallis. "Visual Represenatation of Search Engine Queries and their Results." *1st International Conference on Wbe Information Systems Engineering.* Hong Kong, China, 2000. 352-356.

Havre, S., E. Hetzler, K. Perrine, E. Jurrus, and N. Miller. "Interactive Visualization of Multiple Query Results." *2001 IEEE Symposium on Information Visualization.* San Diego, California, USA, 2001. 105-112.

He, D., and A. Goker. "Detecting Session Boundaries from Web User Logs." *22nd Annual Colloquium of IR Research* . Cambridge, UK, 2000.

Hoeder, O. "Web information retrieval support systems: the future of web search." *the 2008 IEEE/WIC/ACM International Conference on Web Intelligence and Intelligent Agent Technology.* Washington DC, USA,: IEEE, 2008. 29-32.

Hsieh-Yee, I. "Research on Web Search Behavior." *Library and Information Science Research* 23 (2001): 167-185.

Jacso, P. "SAVVY SEARCHING: Clustering Search Results, Part I: Web-wide Search Engines." *Online Information Review* 31, no. 1 (2007): 85-91.

Jansen, B. J., A. Spink, C. Blakely, and S. Koshman. "Defining a session on web search engines." *Journal of the American Society of Information Science* 58, no. 6 (2007): 862-871.

Jing, F., C. Wang, Y. Yao, K. Deng, L. Zhang, and W. Y. Ma. "IGroup: Web Image Search Results Clustering." *14th Annual ACM Internation Conference on Multimedia.* Santa Barbara, CA, USA, 2006. 377-384.

Joho, H., and J. M. Jose. "A comparative study of the effectiveness of search results presentation on the Web." *Lecture Notes in Computer Science, SpringerLink* 3936 (2006): 302-313.

Jones, E., H. Bruce, P. Klasnja, and W. Jones. "I Give Up! Five Factors that Contribute to the Abandonment of Information Management Strategies." *68th Annual Meeting of the American Society for Information Science and Technology* . Columbus,OH, USA: ASIST, 2008. 1-8.

Jones, W., H. Bruce, and S. Dumais. "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.* Zurich, Switzerland, 2003.

K¨aki, M. "Findex: Search Results Categories Help Users When Document Ranking Fails." *SIGCHI Conference on Human Factors in Computing Systems.* Portland, Oregon, 2005. 131-140.

Kaatsen, S., and S. Greenburg. "Integrating Back, History and Bookmarks in Web Browsers." *CHI '01 Extended Abstracts on Human Factors in Computing Systems.* Seattle, WA, USA: ACM, 2001. 379-380.

Karim, J., I. Antonellis, V. Ganapathi, and H. Garcia-Molina. "A Dynamic Navigation Guide for Web Pages." *CHI 2002.* Minneapolis, Minnesota, USA, 2009.

Katifori, A., Halatsis, C., Lepouras, G., Vassilakis, C., Giannopoulou, E. "Ontology visualization methods—a survey." *ACM Computing Surveys (CSUR)* 39, no. 4 (2007): 10.

Kawano, H. "Overview of Mondou Web Search Engine Using Text Mining and Information Visualizing Technologies." *International Conference on Digital Libraries.* Kyoto, Japan, 2000. 234-244.

Kellar, M., C. Watters, and M. Shepherd. "A Field Study Characterizing Web-based Information-Seeking Tasks." *Journal of the Amercian Society for Information Science and Technology* 58, no. 7 (2007): 999-1018.

Kim, K. "Effects of emotion control and task on Web searching behavior." *Information processing and management: an international journal* 44, no. 1 (2008): 373-385.

Knoll, S., A. Hoff, D. Fisher, S. Dumais, and E. Cutrell. "Viewing Personal Data Over Time." *CHI 2009 Workshop on Interacting with Temporal Data.* Boston, MA, USA, 2009. 1-4.

Kobayashi, T., K. Misue, B. Shizuki, and J. Tanaka. "Information Gathering Support Interface by the Overview Presentation of Web Search Results." *the 2006 Asia-Pacific Symposium on Information Visualisation.* Tokyo, Japan, 2006. 103-108.

Kules, B., and R. Capra. "Creating exploratory tasks for a faceted search interface." *the Second Workshop on Human-Computer Interaction (HCIR2008).* Redmond, WA, USA: Microsoft, 2008. 1-4.

Kules, B., W. Wilson, M. C. Schrafel, and B. Sheiderman. "From Keyword Search to Exploration: How Result Visualization Aids Discovery on the Web." School of Electronics and Computer Science, University of Southampton, Southampton, UK, 2008.

Kunz, C., and V. Botch. "Visual Representation and Contextualization of Search Results: List and Matrix Browser." *International Conference on Dublin Core Metadata Applications.* Florence, Italy, 2002. 229-234.

Laycock, J. *Most Search Engine Users Still Naïve.* Search Engine Guide, 2009.

Liu, J.; Cole, M. J.; Liu, C.; Biering, R.; Gwizdka, J.; Belkin, N.; Zhang, J.; Zhang, X. "Search Behaviors in Different Task Types." *10th Annual Joint Conference on Digital Libraries* . Gold Coast, Queensland, Australia, 2010. 69-78.

Mackay, B., and C. Watters. "Exploring Multi-session Web Tasks." *2008 ACM Conference on Human Factors in Computing Systems.* Florence, Italy, 2008. 4273-4278.

Mackay, B., M. Kellar, and C. Watters. "An Evaluation of Landmarks for Re-finding Information on the Web." *2005 ACM Conference on Human Factors in Computing Systems.* Portland, Oregon, USA, 2005. 1609-1612.

Manning, C. D., P. Raghavan, and H. Shutze. *Introduction to Information Retrieval.* New York: Cambridge University Press, 2008.

Marchionini, G. *Information Seeking in Electronic Environments.* New York: Cambridge University Press, 1997.

Mason, J., M. Shepherd, and J. Duffy. "An N-Gram Based Approach to Automatically Identifying Web Page Genre." *42nd Annual Hawaii International Conference on System Sciences.* Waikoloa, Hawaii, USA, 2009. 1-10.

Miller, G. A. "WordNet: An On-line Lexical Database." *International Journal of Lexicography* 3, no. 4 (1990): 285-303.

Moldovan, D., and R. Mihalcea. "Using WordNet and Lexical Operators to Improve Internet Searches." *IEEE Internet Computing* (IEEE) 4, no. 1 (2000): 34-43.

Mountaz, H. "A User Interface Combining Navigation Aids." *11th ACM Conference on Hypertext and Hypermedia.* San Antonio, Texas, USA, 2000. 224-225.

Moyle, M., and A. Cockburn. "The Design and Evaluation of a Flick Gesture for 'back' and 'forward' in Web Browsers." *4th Australian User Interface Conference.* Adelaide, Australia, 2003. 39-46.

Mukherjea, S., and Y. Hara. "Visualizing World Wide Web Search Engine Results." *IEEE International Confernece on Information Visualization.* London, UK, 1999. 400-405.

Murphy, J. "Information-Seeking Habits of Environmental Scientists: A Study of Interdisciplinary Scientists at the Environmental Protection Agency in Research Triangle Park, North Carolina." *Issues in Science and Technology Librarianship*, 2003: No. 38.

Nguyen, T., and J. Zhang. "A Novel Visualization Model for Web Search Result." *IEEE Transactions on Visualization and Computer Graphics* 12, no. 5 (2006): 981-988.

Paulovich, F., R. Pinho, C. B. Botha, A. Heijs, and R. Minghim. "PEx-WEB: Content-based Visualization of Web Search Results." *12th International Conference on Information Visualization (IV'08).* London, UK, 2008. 208-214.

Risden, K., M. P. Czerwinski, T. Munzner, and D. B. Cook. "Initial examination of ease of use for 2 D and 3 D information visualizations of web." *International Journal of Human-Computers Studies* 53, no. 5 (2000): 695-714.

Rivadeneira , W., and B. B. Bederson. "A Study of Search Result CLustering Interfaces: Computing Textual and Zoomable User Interfaces." University of Maryland, College Park, MD, USA, 2003.

Roberts, J. C., N. Boukhelifa, and P. Rodgers. "Visual Depictions of Search Results: Using Glyphs and Coordinated Multiple-Views." *YLEM Journal* 24, no. 2 (2002): 8-10.

Rose, D., and D. Levinson. "Understanding User Goals in Web Search." *13th International Conference on World Wide Web.* New York, NY, USA, 2004. 13-19.

Santini, M. "Interpreting Genre Evolution on the Web." *EACL 2006 Workshop.* Trento, 2006. 32-40.

Santini, M., and S. Sharoff. "Web Genre Benchmark under Construction." *Journal of Language Technology and Computational Lenguistics (JLCL)* 25, no. 1 (2009).

Schraefel, m. c., D. Modjesca, D. Wigdor, and Y. Zhu. "Hunter Gatherer: Within-Web-Page Collection Making." *2002 Conference on Human Factors in Computing Systems.* Minneapolis, Minnesota, USA: ACM, 2002. 498 - 499 .

Sellen, A. J., R. Murphy, and K. L. Shaw. "How Knowledge Workers Use the Web." *the SIGCHI Conference on Human factors in Computing Systems.* Minneapolis, Minnesota, USA: ACM, 2002. 227-234.

Shneiderman, B. "The Eyes Have It: A Task by Data Type Taxonomy for Information Visualizations." *the 1996 IEEE Symposium on Visual Languages.* Boulder, Colorado, USA: IEEE Computer Society, 1996. 336.

Shneiderman, B., D. Feldman, A. Rose, and X. F. Grau. "Visualizing Digital Library Search Results with Categorical and Hierarchical Axes." *5th ACM International Conference on Digital Libraries.* San Antonio, TX, USA, 2000. 57-66.

Shneiderman, Benjamin. *Software Psychology.* 1980.

Spink, A. "Multiple Search Sessions Model of End-User Behaviour: An Exploratory Study." *Journal of the American Society for Information Science* 47, no. 8 (1996): 603-609.

Spink, A., D. Wolfram, M. Jansen, and T. Saracevic. "Searching the Web: the Public and Their Queries." *Journal of the American Society for Information Science and Technology* 52, no. 3 (2001): 226-234.

Srihari, R. K., Z. Zhang, and A. Rao. "Intelligent Indexing and Semantic Retrieval of Multimodal Documents." *ACM Information Retrieval* 2, no. 2-3 (2000): 245-275.

Stubbe, A., C. Ringlstetter, T. Zheng, and R. Goeble. "Incremental Genre Classification." *Colloquim Hel in Conjunction with Corpus Linguistics.* Birmingham, UK, 2007.

Sugiyama, K., K. Hatano, and M. Yoshikawa. "Adaptive web search based on user profile constructed without any effort from users." *the 13th international conference on World Wide Web.* New York, NY, USA, 2004. 675-684.

Sutcliffe, A. G., and U. Patel. "3D or not 3D: is it nobler in the mind?" *the BCS Human Computer Interaction Conference on Peopleand Computers XI.* London, UK, 1996. 79-94.

Suvanaphen, E., and J. C. Roberts. "Textual Difference Visualization of Multiple Search Results Utilizing Detail in Context." *Theory and Practice of Computer Graphics Conference.* Bournemouth, UK, 2004. 2-8.

Tao, X., and Y. Li. "Concept-Based, Personalized Web Information Gathering: A Survey." *3rd International Conference on Knowledge Science, Engineering, and Management.* Vienna, Austria, 2009. 215-228.

Taucher, L., and S. Greenberg. "How People Revisit Web Pages: Empirical Findings and Implications for the Design of History Systems." *International Journal of Human Computer Studies* 47, no. 1 (1997): 97-138.

Teevan, J. "How People Recall, Recognize and Reuse Search Results." *ACM Transactions on Information Systems special issue on Keeping, Refinding, and Sharing Personal Information* 26, no. 4 (2008).

Teevan, J., C. Alvarado, M. S. Ackerman, and D. R. Karger. "The Perfect Search Engine is not Enough: A Study of Orienteering Behaviour in Directed Search." *2004 Conference on Human Factors in Computing Systems.* Vienna, Austria, 2004. 415-422.

Teevan, J., et al. "Visual Snippets: Summarizing Web Pages for Search and Revisitation." *27th International Conference on Human Factors in Computing Systems.* Boston, MA: ACM, 2009. 2023-2032.

Tilsner , M., O. Hoeber, and A. Fiech. "CubanSea: Cluster-Based Visualization of Search Results." *IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology.* Milano, Italy, 2009. 108-112.

Turetken, O., and R. Sharda. "Clustering-based Visual Interfaces for Presentation of Web Search Results: An Imperical Investigation." *Information Systems Frontier* 7, no. 3 (2005): 273-297.

Tyler, S. K., and J. Teevan. "Large scale query log analysis of re-finding." *the 3rd ACM International Conference on Web Search and Data Mining.* Santa Cruz, CA, USA, 2010.

Wang, A. G., J. Jiao, and W. Fan. "Searching for Authoritative Documents in Knowledge-based Communities." *13th International Conference on Information Systems.* Phoenix, AZ, USA, 2009. Paper N. 109.

Wilson, T. D. "Models in Information Behaviour Research." *Journal of Documentation* 55, no. 3 (1999): 249-270.

Wilson, T. D., and C. Walsh. "Information Behaviour: an Interdisciplinary Perspective." British Library Research and Innovation Report, University of Sheffield, Department of Information Studies, Sheffield, UK, 1996.

Wiza, W., K. Walczak, and W. Cellary. "Periscope: a System for Adaptive 3D Visualization of Search Results." *9th International Conference on 3D Web Technology.* Monterey, CA, USA, 2004. 29-40.

Woodruff, A., A. Faulring, R. Rosenholtz, J. Morssion, and P. Pirolli. "Using Thumbnails to Search the Web." *SIGCHI Conference on Human Factors in Computing Systems.* Seattle, Washington, USA, 2001. 198-205.

Woodruff, A., Rosenholtz, R., Morrison, J., Faulring, A.,Pirolli, P. "A. Woodruff, R. Rosenholtz, J. Morrison, A Comparison of the Use of Text Summaries, Plain Thumbnails, and Enhanced Thumbnails for Web Search Tasks." *Journal of the American Society for Information Science and Technology* 53, no. 2 (2002): 172-185.

Yamada, S., and H. Kawano. "Information Gathering and Searching Approaches on the Web." *New Generation Computing* 19, no. 2 (2009): 195-208.

Yamaguchi , T., H. Hattori, T. Ito, and T. Shintani. "On a Web Browsing Support System with 3D Visualization." *13th International WWW Conference on Alternate Track Papers and Posters.* New York, NY, USA, 2004. 316-317.

Zaina, C., and M. C. Baranauskas. "Revealing Relationships in Search Engine Results." *2005 Latin American Conference on Human-Computer Interaction.* Cuernavaca, Mexico, 2005. 120-127.

Zamir, O., and O. Etzioni. "Grouper: a Dynamic Clustering Interface to Web Search Results." *8th International Conference on World Wide Web.* Toronto, ON, Canada, 1999. 1361-1374.

Zelberstein, S., and V. Lesser. "Intelligent Information Gathering Using Decision Models." Computer Science Department, University of Massachusetts, Boston, Massachusetts, 1996.

Zhuang, Z., and S. Cuserzan. "Re-ranking Search Results Using Query Logs." *15th ACM International Conference on Information and Knowledge Management.* Arlington, Virginia, USA, 2006. 860-861.

Zitouni, H., S. Sevil, D. Ozkan, and P. Duygulu. "Re-ranking of Web Image Search Results Using a Graph Algorithm." *19th International Conference on Pattern Recognition.* Tampa, FL, USA, 2008. 1-4.

APPENDIX A Pre-study Questionnaire

1. What is your age? Please circle one of the following:

   18-24 ***25-35*** ***>35***

2. How long have you been using the web to search for information? ………… **Years**

3. What tools (in addition to the web browser) do you usually utilize to put together pieces of information you are gathering on the web for a report or a project you are trying to complete such as when you are writing an essay, a report, or a school project?

   ……………………………………………..……………………………………

   ………………………………………………...………………………………

4. What best describes your experience of the use of web search engines and other tools to find web information?

   very well-experienced ***experienced*** ***not sure*** ***a little experienced*** ***inexperienced***

5. How often do you use the web to gather information such as for a report or a project?

   always ***often*** ***sometimes*** ***rarely*** ***never***

6. When you gather web information for a project or to write a report, how difficult do you find organizing such information, especially when more than one source of information is involved or when you need more than one session to gather all the required information? Please circle one of the following:

difficult ***somewhat difficult*** ***neither difficult nor easy*** ***somewhat easy*** ***easy***

If any, why do you think there is difficulty in organizing your information?

……………………………………………………………………………..……………

……………………………………………………..………………………

7. How often do you need more than one session to finish gathering information on the web for your task or project?

   always ***often*** ***sometimes*** ***rarely*** ***never***

8. Please indicate tools and systems you usually use to manage web information you are gathering.

   ……………………………………………………………………………………

   ……………………………………………………………………………………

   ……………………………………………………………………………………

9. Please describe difficulties you have had when you try to gather information on the web to complete a task such as a school report or a project.

   ………………………………………………..…………..……………………

   ……………………………………………………………………………………

   ……………………………………………………………………………………

APPENDIX B POST-STUDY QUESTIONNAIRE

1. Was the task description clear? (Yes/No)

2. How would you rate the task taking into account your progress and achievement of the task goal?

   Difficult   somewhat difficult   neither difficult nor easy   somewhat easy   easy

3. Did you complete the task as described?

   a. I completed the task.

   b. Some parts of the required information are still missing.

   c. I completed all the requirements but I am not sure they are correct/relevant.

   d. I could not complete the whole task because (Please specify):

   ………………………………………………………………………………

   ………………………………………………………………………………

   e. Other, please specify:

   ………………………………………………………………………………

   ………………………………………………………………………………

4. How confident are you that you satisfied the requirements of the task?

   Very confident   Confident   Not sure   Not confident   Not confident at all

5. Please describe any difficulties you encountered during the task.

   ………………………………………………………………………………

   ………………………………………………………………………………

   ……………………………………………………………………………

6. What kind of tools would have helped you complete the task more effectively and quickly?

…………………………………………………………………………………

…………………………………………………………………………………

……………………………………………………………………………..……

7. Which of the following tools and strategies you used during the task to organizing and manage your information? Please circle tools you used

    a. Text editor such as MSWord. To do what?

       …………………………………………………………………………

       …………………………………………………………………………

       …………………………………………………………………………

    b. Emailing. For what reasons?

       …...

       …………………………………………………………………………

       …………………………………………………………………………

    c. Store information on the computer:

        i. Temporarily

        ii. Permanently

        iii. Both

       Reasons for storing information:

       ….……………………….……………………………………….………

       …………………………………………………………………………

       …………………………………………………………………………

    d. Copy information from documents. Why?

       ……………………………………………………..……………….

……………….…………………………………………………………………

…………………………………………………………………………………

…………………………………………………………………………………

e. Store links of web pages. Why?

……………………………………………………….……………...……

………...……………………………………………………………………

…………………………………………………………………………………

…………………………………………………………………………………

f. Print out information? Why?

……………………………………………...………………………………

……………………………………………...………………………………

……………………………………………...………………………………

g. Other, please specify:

…………………………………………………………………………………

…………………………………………………………………………………

…………………………………………………………………………………

…………………………………………………………………………………

…………………………………………………………………………………

APPENDIX C INTERVIEW

The interview sought reasons for committing to certain behaviour while the user gathers and organizes information for the task. Why users use certain tools and why they follow particular orders were also discussed during the interview. Items selected for the interview depended on the task accomplishment process, the user, and the results gathered for the task. Some of the interview items involved questions such as:

1- Why did you use ***a certain*** tool?

2- How often do you perform a task ***this way***?

3- Have you ever had the same experience or worked with similar tasks?

4- Did you find what you are looking for? Why/Why not?

5- What was ***hard/easy*** about the task?

6- What tools do you think were missing? How would they have helped you with your task?

7- Are you happy with the information you gathered in the task? Did you get the necessary information? ***Did you manage to put together (organize) the information as required in the task in order to present the results?***

8- Other questions that depended on the situation such a tool used, a strategy used, or something the user indicated.

APPENDIX D PRE-STUDY QUESTIONNAIRE

1. What is your ID?

2. Are you

○ Male

○ Female

3. What is your age?

○ 18-22

○ 23-30

○ >30

4. Do you use the web to gather information for reports, papers, booking trips, and the like?

○ Yes

○ No

5. Please select the tools and features that you use during gathering and collecting information from the web:

□

□ Text editor such as Word

□ Local bookmarking

□ Online bookmarking

□ Session saving

Other (please specify)

6. Which of the following do you usually have difficulties with?

□ saving pages

□ saving information together

□ creating bookmarks

□ retrieving bookmarks

□ re-locating a task on which you worked in previous sessions

□ saving sessions

□ editing and browsing for information

□ searching and browsing for information

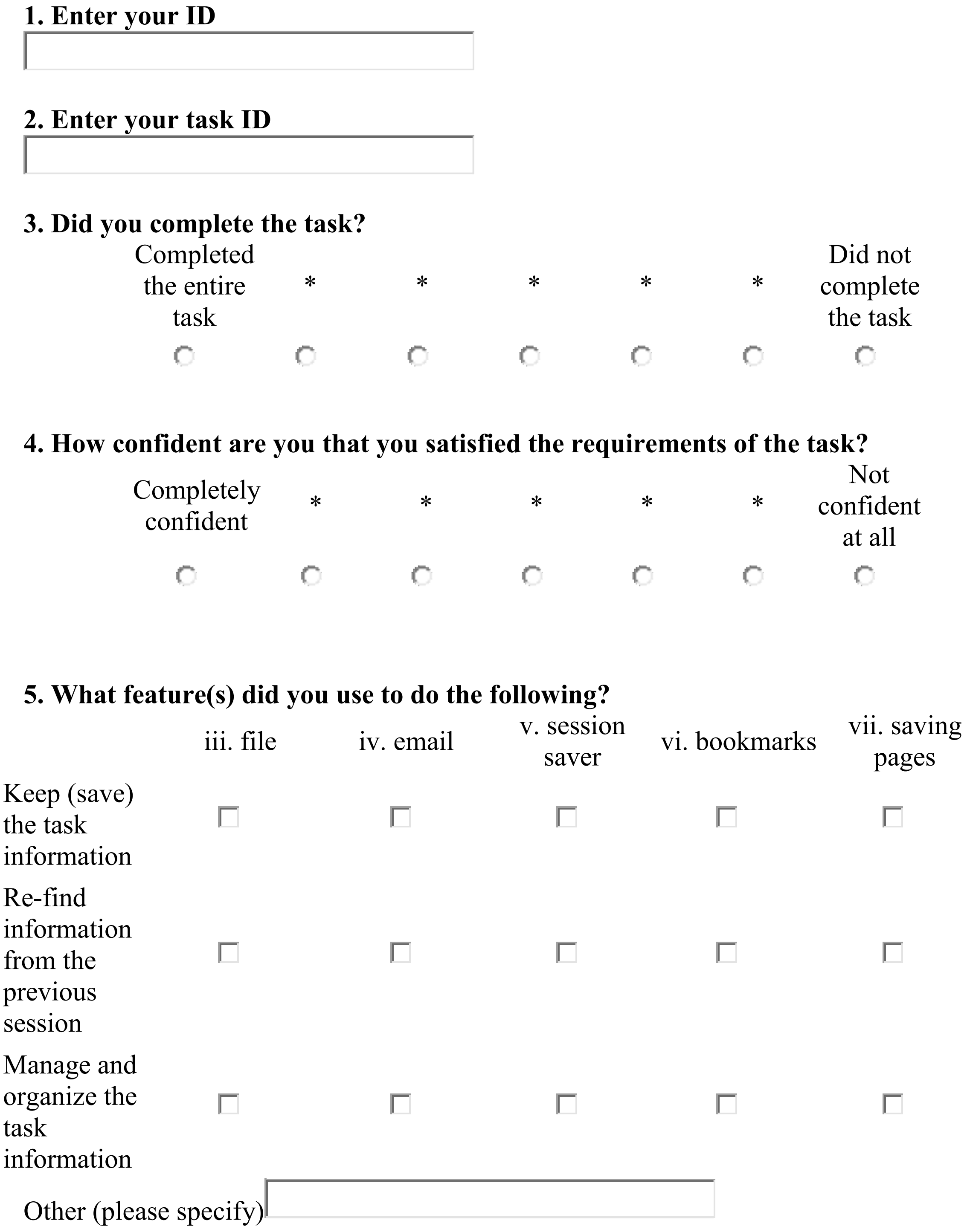
□ searching and managing information for a task

□ saving open tabs together

□ relocating all open sources in previous sessions

Other (please specify)

APPENDIX E Post-Task Questionnaire (Browser)



1. Enter your ID

2. Enter your task ID

3. Did you complete the task?

Completed the entire task	*	*	*	*	*	Did not complete the task
○	○	○	○	○	○	○

4. How confident are you that you satisfied the requirements of the task?

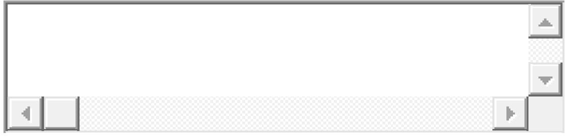
Completely confident	*	*	*	*	*	Not confident at all
○	○	○	○	○	○	○

5. What feature(s) did you use to do the following?

	iii. file	iv. email	v. session saver	vi. bookmarks	vii. saving pages
Keep (save) the task information	□	□	□	□	□
Re-find information from the previous session	□	□	□	□	□
Manage and organize the task information	□	□	□	□	□

Other (please specify)

6. What other features would have helped you with completing the task?



APPENDIX F Post-Task Questionnaire (WIGI)

1. Enter your ID

2. Enter your task ID

3. Did you complete the task?

Completed the entire task	*	*	*	*	*	Did not complete the task
○	○	○	○	○	○	○

*

4. How confident are you in the information you gathered for the task?

Completely Confident	*	*	*	*	*	Not confident at all
○	○	○	○	○	○	○

*

5. How effective was the following:

	Very effective	*	*	*	*	*	NOT effective at all
thumbnails for tracking references.	○	○	○	○	○	○	○
ability to edit and format information along with browsing and searching.	○	○	○	○	○	○	○
ability to embed references while editing information.	○	○	○	○	○	○	○
ability to keep all the task information as one unit.	○	○	○	○	○	○	○
ability to re-find all the task information as one unit.	○	○	○	○	○	○	○
The use of WIGI as a whole.	○	○	○	○	○	○	○

6. What feature(s) did you use to do the following?

	i. the 'save gathered' feature	ii. the 'restore task' feature	iii. file	iv. email	v. session saver	vi. embedded editor
Keep (save) the task information	□	□	□	□	□	□
Re-find information from the previous session	□	□	□	□	□	□
Manage and organize the task information	□	□	□	□	□	□

Other (please specify)

*

7. What other features would have helped you with completing the task?