

A Goal-based Classification of Web Information Tasks

Melanie Kellar

Faculty of Computer Science, Dalhousie University, Halifax, Nova Scotia,
Canada melanie@cs.dal.ca

Carolyn Watters

Faculty of Computer Science, Dalhousie University, Halifax, Nova Scotia,
Canada watters@cs.dal.ca

Michael Shepherd

Faculty of Computer Science, Dalhousie University, Halifax, Nova Scotia,
Canada shepherd@cs.dal.ca

While researchers have been studying user activity on the Web since its inception, there remains a lack of understanding of the high level tasks in which users engage on the Web. We have recently conducted a field study in which participants were asked to annotate all web usage with a task description and categorization. Based on our analysis of participants' recorded tasks during the field study, as well as previous research, we have developed a goal-based classification of information tasks which describes user activities on the Web.

INTRODUCTION

Over the past fifteen years the Web has drastically changed how we acquire and circulate information. We now conduct our research using search engines and online library portals, read the daily news and our favourite comic online, communicate with others increasingly through email and blogs, and have become accomplished fact checkers thanks to Google. However, researchers still lack a solid understanding of the types of activities and tasks in which users engage on the Web.

There are several reasons for this lack of understanding. First, the Web is a moving target and is continually changing and evolving. For example, the typical user has changed substantially since the early 1990s when the average web user was a young, technically inclined male (Hawkey and Inkpen, 2005b). Also, the Web now supports a much wider

range of activities and uses. Examples include the increase in web-based email; new sophisticated web-based travel and map applications; and the popularity of online support and blog communities. Therefore, models and frameworks developed in the 1990s need to be continually validated against the current state of the Web.

Second, collecting rich and detailed user data on the Web can be very difficult from a methodological standpoint. While research conducted in the field provides a more realistic picture of users' natural behaviour on the Web, it is difficult to carry out because there is a lack of appropriate tools for collecting both contextual information (e.g., user task) and detailed web browser interactions (Fenstermacher and Ginsburg, 2003; Hawkey and Inkpen, 2005b). Alternatively, laboratory research is somewhat easier to conduct from a methodological standpoint but can impede a user's natural behaviour as the tasks are often contrived and users do not typically have access to their usual web browsers and tools.

Third, user behaviour on the Web is varied and complex. The diverse and dynamic nature of the Web means that users are engaging in a wide variety of tasks and activities and this range of activity varies across users and populations. There is evidence to suggest that users' search behaviour may differ between home and work environments (Rieh, 2003). Individual differences between users may play a role in users' navigation styles (Herder and Juvina, 2004). Cognitive differences (Ford, Wilson, Foster, Ellis and Spink, 2000), as well as domain knowledge and web experience (Hölscher and Strube, 2000), may play a role in a user's information seeking behaviour. All of these factors make it difficult to compare the behaviour of participants and to generalize results to communities of web users.

Researchers who want to characterize user activity on the Web often look to models of information seeking (Ellis, 1989; Marchionini, 1995; Choo, Detlor and Turnbull, 2000). While these models provide good characterizations of users' information seeking activities, there are a number of activities undertaken by users on the Web that are not described by these models. There are also variations among the different models and theories that can make it difficult to choose an appropriate characterization.

In this paper we have conducted an in-depth exploration of the higher-level tasks in which users engage in on the Web, using data collected during a week-long field study (Kellar, Watters and Shepherd, 2006). Throughout the study, participants used a custom built web browser that logged all of their interactions on the Web (including the use of web browser navigation mechanisms, browser functions, and URLs visited). Participants also annotated all web usage with task information using an electronic diary format. Annotations consisted of a short task description as well as a categorization of the task according to

the following schema: Fact Finding, Information Gathering, Browsing, Transactions, and Other.

The primary contribution of this paper is a new high-level classification of users' activities on the Web according to their web information goals. This classification is based on observations during the field study as well as earlier models and frameworks (Choo, Detlor and Turnbull, 2000; Morrison, Pirolli and Card, 2001; Sellen, Murphy and Shaw, 2002). In the next section, we provide an overview of the related work. We then describe the development of the task categorization used in this study, followed by a description of the methodology used during the field study. The results section presents our observations of participants' web activities during the field study. Finally, we introduce the web information task classification and conclude with a summary of our future work.

RELATED WORK

In this section we review relevant theories and models of information seeking as well as studies of user behaviour on the Web.

Theories and Models of Information Seeking

A large number of information seeking models and theories exist that attempt to help us understand how users satisfy their information needs. Much of users' activity on the Web is related to a need for information; therefore, models of information seeking behaviour have been used to classify users' web-based tasks. We focus on a subset of the common user centred models of information seeking.

Belkin's (1980) model of information seeking is based on anomalous states of knowledge (ASK), which occur when the information seeking problem is not well defined and users must iterate on their queries and requests before they can be adequately expressed. Ellis (1989) initially developed a behavioural model of information seeking that consisted of six activities: starting, chaining, browsing, differentiating, monitoring, extracting. Two more activities, verifying and ending, were later added by Ellis, Cox, and Hall (1993). Meho and Tibbo (2003) further extended Ellis' model to support web-based information seeking through three additional activities: accessing, networking, and information managing.

Kuhlthau's (1991) model of the information search process is in some ways similar to Ellis' (1989) model of information seeking but also incorporates feelings, thoughts, and actions. The model consists of six stages of information seeking: initiation, selection, exploration, formulation, collection, and presentation.

Marchionini's (1995) model of information seeking describes a series of subprocesses. The subprocesses consist of recognizing the information problem, understanding the problem, choosing a search system, formulating a query, executing the search, examining the results, extracting the relevant information, and deciding to stop/reflect/iterate upon the search process. This model was developed in the context of electronic document environments.

Wilson and Walsh's (1996) model of information behaviour differs from many of the previous models by suggesting more high-level information seeking search processes: passive attention, passive search, active search, and ongoing search. Passive attention occurs when information is obtained without being actively sought such as while listening to the radio or television. Passive search is the serendipitous acquisition of information through search. Active search occurs when information is actively being sought through explicit searches. Ongoing search occurs when occasional searching is carried out to expand or update previously found information.

While the previously presented models of information seeking have been very useful in understanding information seeking behaviour, they cannot be used to characterize all tasks in which users engage on the Web. Examples include serendipitous browsing tasks,

fact re-checking, news reading, or online transactions (such as email or banking).

Web-based Information Seeking

Other studies have examined general user behaviour on the Web. In one of the first studies of web usage, Catledge and Pitkow (1995) classified user strategies into three categories: serendipitous, general purpose, and searcher. Pitkow and Kehoe (1996) reported five main uses of the Web from the fourth GVV WWW survey: browsing, entertainment, work, shopping, and other uses. They also noted that the activities had remained fairly consistent since the second study.

Table 1

	Choo et al.(2000)	Morrison et al.(2001)	Sellen et al.(2002)	Rozanski et al.(2002)
1	Informal search	Find	Finding	Just The Facts/Quickies
2	Formal Search	Collect	Information Gathering	Information Please/SingleMission
3	Undirected Viewing	Explore	Browsing	Surfing/Loitering
4	ConditionedViewing	Monitoring	N/A	Do It Again
5	N/A	N/A	TransactingCommunicatingHousekeeping	N/A

Choo, Detlor, and Turnbull (2000) studied critical incidents of information seeking on the Web among 34 knowledge workers. Using interviews, questionnaires, and data logging over a two week period, significant episodes of information seeking were characterized as undirected viewing, conditioned viewing, informal search, and formal search. This characterization was partially based upon both Ellis' (1989) model and Wilson and Walsh's (1996) model of information seeking.

Morrison, Pirulli, and Card (2001) studied significant web actions through 2188 responses to the 10th WWW user survey. Participants were asked to describe a recent episode in which they found information on the Web that led to a significant decision or action. The participants reported four main goals: collect, find, explore, and monitor.

Sellen, Murphy, and Shaw (2002) studied the web activities of 24 knowledge workers over two days. Participants were interviewed in front of their of web history at the end of the

second day and described the different activities in which they engaged. Activities were classified into six main categories: finding, information gathering, browsing, transacting, communicating, and housekeeping.

Finally, Rozanski, Bollman, and Lipman (2001) reported seven main web usage occasions: quickies, just the facts, single mission, do it again, loitering, information please, and surfing. This work was conducted from a commercial standpoint since the focus of their work was for marketing purposes.

Although these studies differed in methodology and research goals, there are strong similarities among the resultant categorizations (shown in Table 1). The first is the short answer or informal search, including fact finding and simple lookup. In this category the goal of the user is to retrieve some short, specific information, possibly on one page. The second category, the formal search, is the more traditional bibliographic search in which the user's goal is to collect enough information on a topic to write a report or make a decision. This may require multiple pages and overlapping data for confirmation or alternate views on the topic. The third category is the ludic notion of browsing, where the user is engaged in serendipitous information seeking. The fourth category is monitoring, which includes repeated visits to one or more web pages to monitor or check for dynamic information. As can be seen in Table 1, monitoring is not always included as a distinct information seeking task. The fifth category consists of the remaining web tasks studied by Sellen, Murphy and Shaw (2002) which consist of non information seeking tasks such as transacting (e.g., online transactions), communicating (e.g., chat rooms and discussion boards), and housekeeping (e.g., maintaining web pages).

Based on these previous works, we developed an initial task categorization consisting of the following tasks: Fact Finding, Information Gathering, Browsing, and Monitoring. However, before beginning the week long field study we needed to verify that the categories reflected most of the tasks in which users now engage on the Web while at the same time remaining easy to understand and relatively distinct. In the next section, we describe how we iterated on this initial task categorization through a pilot study and focus group

TASK CATEGORIZATIONS

This section describes the methodologies and findings from the pilot study and focus group.

Pilot Study

In preparation for the field study, a four day pilot was conducted with six participants who were all recruited from within our research lab at Dalhousie University. Participants were asked to use a custom web browser for all their web usage during the pilot, which logged all interactions with the browser (including URLs visited). Participants were also asked to categorize their web usage according to the following four categories: Fact Finding, Information Gathering, Monitoring, and Browsing. A fifth task of Other was provided for all other tasks that did not fit within the given schema. A fifteen minute training session was provided for all participants before beginning the pilot. Upon completion of the pilot study, participants completed a post-session questionnaire to explore their attitudes towards the logging software and the task categorization.

The goal of the pilot study was to evaluate how well participants were able to categorize their web usage. Overall, we found that participants struggled with the task of Monitoring. When revisiting websites, participants found it difficult to judge whether they were actually engaging in Monitoring. They found it difficult to distinguish Monitoring from “re-Fact Finding” or “re-Browsing”. One example was reading online comics. A participant was unsure whether repeatedly reading the same comic strip was Browsing or Monitoring. It seems reasonable that the task could be categorized as either. In addition, we had not anticipated the high level of web-based email and other online transactions, which participants had tagged as Other. Therefore, a focus group was planned to further refine the task categorizations.

Focus Group

Ten participants from the Faculty of Computer Science (students and faculty) at Dalhousie University took part in a one hour informal focus group designed to help finalize an appropriate categorization of tasks for the field study. Participants had backgrounds in Web Behaviour, Human-Computer Interaction, and Information Science.

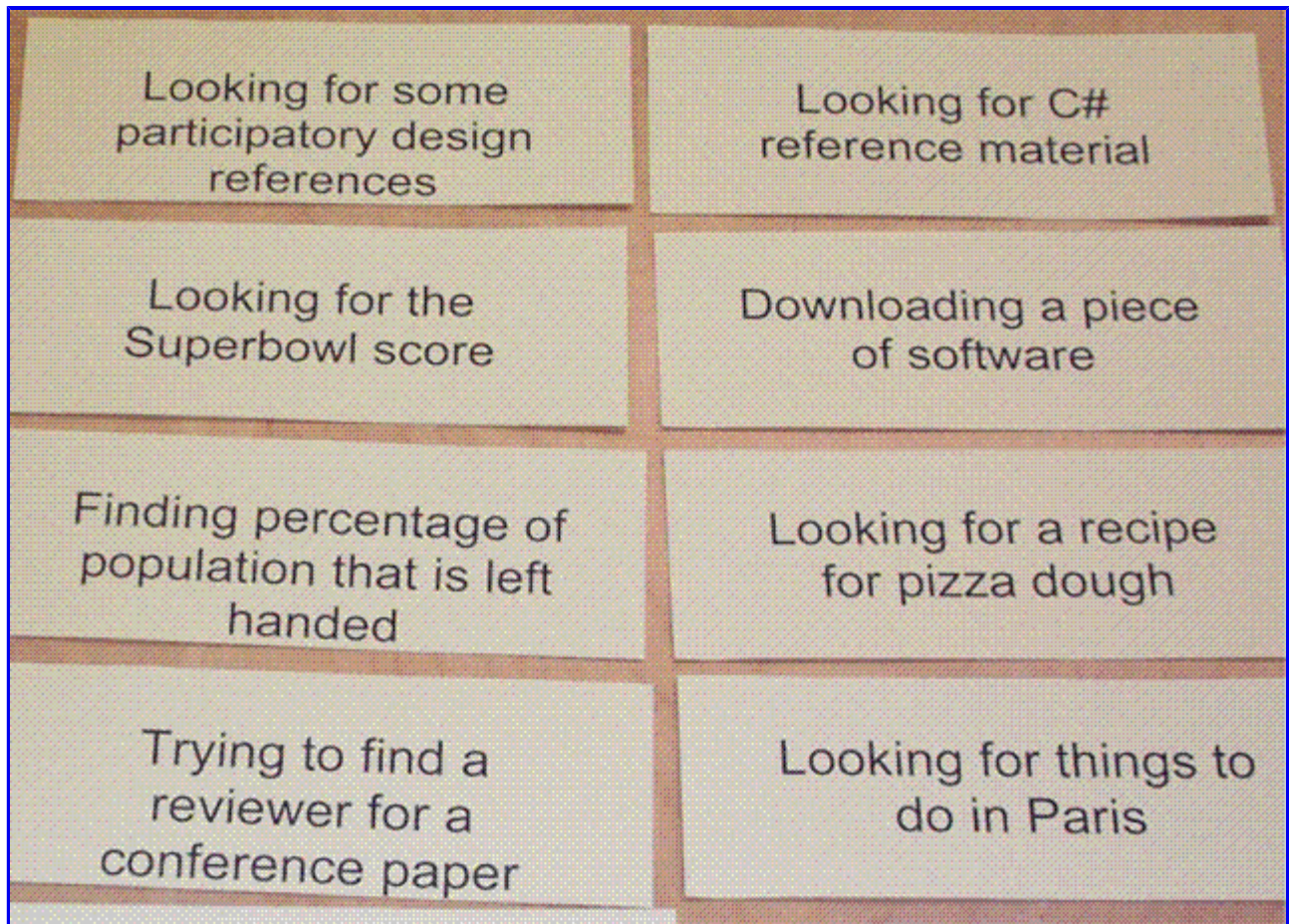


Figure 1. Cue cards, each containing a task description, were used to refine the tasks.

We selected forty task descriptions from the set of task descriptions collected during the pilot study to use during the focus group. Each task description was printed onto an index card and spread out on a large table (as shown in Figure 1). The participants were instructed to work together to organize the index cards into distinct groups based on the goal of the task printed on the card. None of the focus group participants were informed of the categories used in the pilot study or in previous literature.

During the hour long focus group, participants re-arranged the task groupings several times. The content and number of categories fluctuated continually during the course of the session. After much discussion among the participants, the categories began to stabilize and six final categories emerged (shown in Table 2). We labelled the categorizations produced by the focus group participants as: Looking for Specific Information, Information Gathering, Transactions & Communication, Routine & Hobby, Passing Time & Entertainment, and Monitoring.

The task categories that evolved out of this focus group were in fact very similar to the tasks reported in the literature. Based on the findings of our pilot, however, we suspected

that Monitoring may actually occur within several tasks and may be very difficult for participants to identify. Therefore, we decided to eliminate Monitoring from the categories and instead study Monitoring informally using the task descriptions provided by the participants. We also merged the categories Passing Time & Entertainment and Routine & Hobby into a single category (Browsing) as it was difficult to clearly articulate the distinction between these two categories as they are both serendipitous in nature and lack specific goals. The resulting task categories, shown in Figure 2, were: Fact Finding, Browsing, Information Gathering, Transactions. A category of Other was again provided for the tasks that did not fit within the given schema. For the ensuing field study, the following task descriptions were provided to participants:

Table 2

Task	Examples
<i>Looking for Specific Information</i>	Location of a conference workshop Finding percentage of the population that is left handed
<i>Passing Time & Entertainment</i>	Random Surfing Just browsing EBay
<i>Transactions & Communication</i>	Checking my email Online banking
<i>Information Gathering</i>	Trying to find a reviewer to review a conference paper Looking for references on a topic
<i>Routine & Hobby</i>	Reading my favourite comic Reading blogs
<i>Monitoring</i>	Checking to see if a project page is up to date Looking up the prices of my stocks

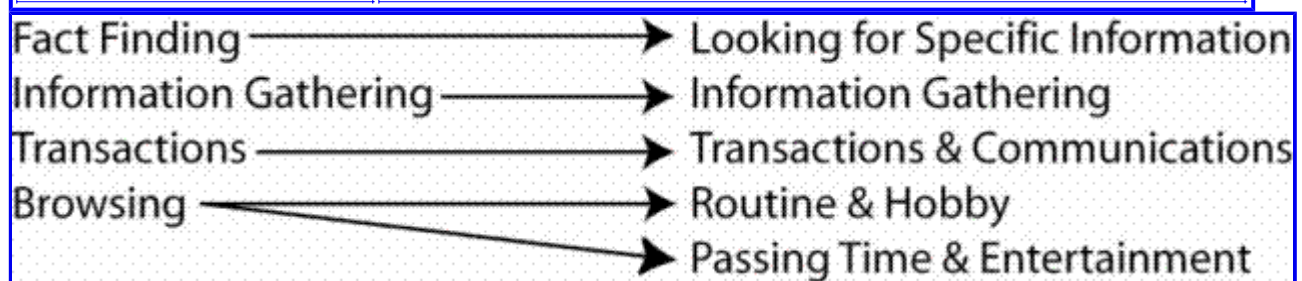


Figure 2

Fact Finding: A task in which you are looking for specific facts or pieces of information. These are usually short lived tasks that are completed over a single session because either you find the answer or you do not. Examples include looking for tomorrow's weather, a pizza dough recipe, or printer drivers for your printer.

Information Gathering: A task that involves the collection of information, often from multiple sources. This type of task can take place over a single day or may stretch out over several days. Unlike Fact Finding, you do not always know when you have completed the task and there is no one specific answer. Examples include building a bibliography for a research paper, researching different car models when buying a new car, or planning an upcoming vacation.

Browsing: A serendipitous task where you may visit web pages with no specific goal in mind. You may allow yourself to take part for a pre-determined period of time (e.g., I have 20 minutes before my meeting). This type of task is your classic “web browsing”, with no specific goal in mind other than entertainment or to “see what’s new”. Sometimes this is done as part of a daily routine. Examples include reading the news, your favourite comic, or a friend’s blog.

Transactions: Tasks in which you are performing an online action. Often, a username/password is associated with the transaction. Examples include web-based email, banking, or posting to a message board.

Other: Tasks which do not fit within any of the predefined categories.

METHODOLOGY

This section presents the methodological details of the field study.

Sample Population and Procedure

Twenty-one university students from Dalhousie University took part in a one week field study in March, 2005. Although 23 participants were recruited, only data for 21 participants was analyzed. One of the original participants did not finish the study and another participant’s data was unusable because the task descriptions were incomplete and inconsistent. Laptop users were targeted so that we could capture most of their web usage on a single machine and because it facilitated installation of the custom software. Also, since the web browser used during the study was a clone of Microsoft Internet Explorer (IE), participants were required to be current users of IE.

The academic background of the participants was divided among Computer Science (11/21), Health Informatics (2/21), Business (4/21), Economics (2/21), Kinesiology (1/21), and Arts (1/21). Participants were also from both the graduate and undergraduate

communities: Computer Science (7 grad/4 undergrad), Health Informatics (2 grad), Business (4 grad), Economics (2 grad), Kinesiology (1 undergrad), and Arts (1 undergrad). The median age group category of the participants was 20-29 and the gender was almost evenly split with 11 males and 10 female participants. The median category of web usage reported by the participants was between 30-39 hours of web usage a week. Although Computer Science students are typically considered to be more highly technical, all participants were experienced web users.

On the first day of the study, each participant met with the researcher administering the study for a one hour session in which a custom web browser and electronic diary were installed on the participant's laptop. The custom web browser was configured with the same settings as the participant used in IE, such as auto-complete, the bookmarks toolbar and the Google toolbar. A demographic and web browser tools inventory questionnaires were administered at this time. The researcher then carefully described the different task categories and explained how to use both methods to record the task information. Participants then took part in a short training exercise in which they were required to complete several short information seeking tasks using both the toolbar and task diary methods of providing task information. Finally, participants were given printouts of the task definitions (which were also available online) and instructions for the study tools. After a one week period, participants returned to meet with the same researcher. The software was uninstalled from the participant's laptop and all logging data was copied on a backup disk and then deleted. Participants completed a final post-study questionnaire and were paid \$25 for their participation in the study.

Data Collection

During the course of the study we captured participants' web usage, task information, and questionnaire data. Capturing a detailed picture of participants' web usage can be difficult. We explored several commercial and academic software logging tools, however none of the standard logging solutions met all our needs. Therefore, we built a custom web browser (shown in Figure 3a) in C# using the browser control function provided by Microsoft .Net. The custom browser mimicked the appearance of IE, the functionality (including a Google search toolbar), and logged all usage. All history and bookmark files were shared between IE and the custom

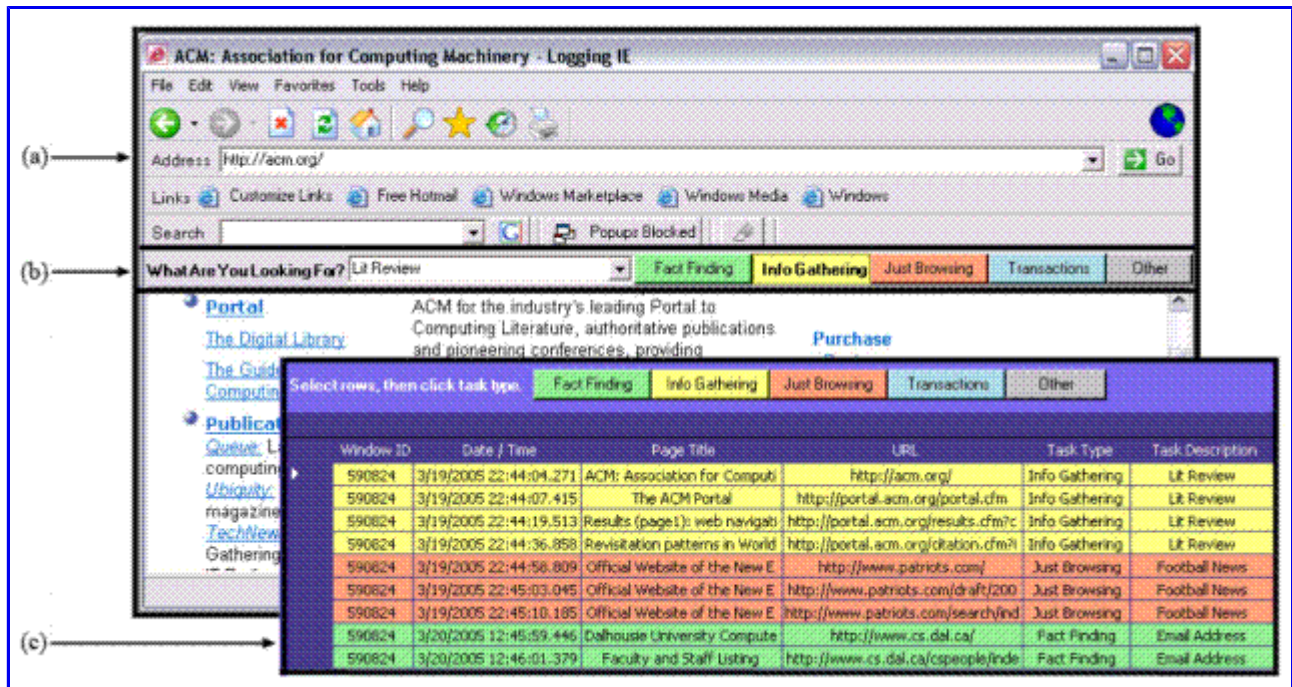


Figure 3. The custom web browser (a) was built to mimic IE and provided a task toolbar (b) for participants to record their task information in real-time. Participations could also use the task diary (c) to record their task information at the end of each day.

web browser, meaning that on the first day of the study, participants had full access to their recent history data and did not need to create new (or import) bookmarks. It was of critical importance that participants used a familiar web browser, with all of their usual tools and this was reflected in the design of the custom web browser.

Participants were asked to categorize all web usage according to the task categories described in the previous section, as well as to provide a short textual description of the task (e.g., “reading the news”, “looking for an email address”). Based on the results of the pilot study, where user preference was evenly split, participants were given the option to provide their task information in real-time using a toolbar built into the custom web browser (shown in Figure 3b) or at the end of the day using a task diary (shown in Figure 3c). Alternatively, participants could use a combination of both tools.

Participants who preferred the toolbar method were instructed to fill in task information at the beginning of a new task. Participants who preferred to use the task diary to assign task information were instructed to do so at the end of each day. The task diary, similar to the approach used by Hawkey and Inkpen (2005a) to collect privacy information, allowed participants to assign task information to multiple URLs at once. In an effort to encourage users to work on the Web as they normally would, all participants could use the task diary to delete URLs in which they were uncomfortable sharing with the researchers involved in

the study. Regardless of the method used to collect task information, each URL visited was associated with a task categorization and description. This information was recorded in a log file in the following format: window id, date & time, page title, URL, task categorization, and task description. A more detailed discussion of the data collection techniques can be found in (Kellar, Watters and Shepherd, 2005).

Participants completed three separate questionnaires over the course of the study. During the pre-study session, a demographic questionnaire was used to collect participants' demographic information and current web usage. An inventory questionnaire of the web browser tools used was also completed by participants during the pre-study session. Upon completion of the study, participants completed a post-study questionnaire which examined any difficulties they encountered during the study.

Upon completion of the study and before analysis of the data, a single researcher manually reviewed all participants' data. Occasionally, we encountered instances where the task information did not appear to match the URLs recorded. In many cases, participants had forgotten to update their task information as they switched to a new task. Only in cases where the behaviour was habitual and obvious did the researcher alter the task information. In all other cases, the participants were contacted in order to clarify the task information.

RESULTS

Earlier analysis of data from this field study examined differences in the use of web browser navigation mechanisms across information seeking task sessions (Kellar, Watters and Shepherd, 2006). The focus of this paper is the exploration of the types of tasks in which users engaged in during the week long field study

Task sessions were defined as a period of continuous usage, annotated with the same task information, with no break greater than 25.5 minutes, similar to the approach used by Catledge and Pitkow (1995). In the case of Transactions, a new task was identified either using the 25.5 minute lapse in activity or an explicit session logout indicated by the existence of the "logout" string in a Transaction URL (e.g., <http://www.mail.yahoo.com/logout>). Overall, participants recorded 1192 task sessions involving 13,498 pages over the week long study. The mean number of tasks completed per participant was 56.8 (median = 52, SD = 31.97) with a range of 16 to 140 tasks. The breakdown of all tasks is shown in Figure 4. An overview of the proportion of repeated tasks, as well as the most commonly repeated tasks, is shown in Table 3.

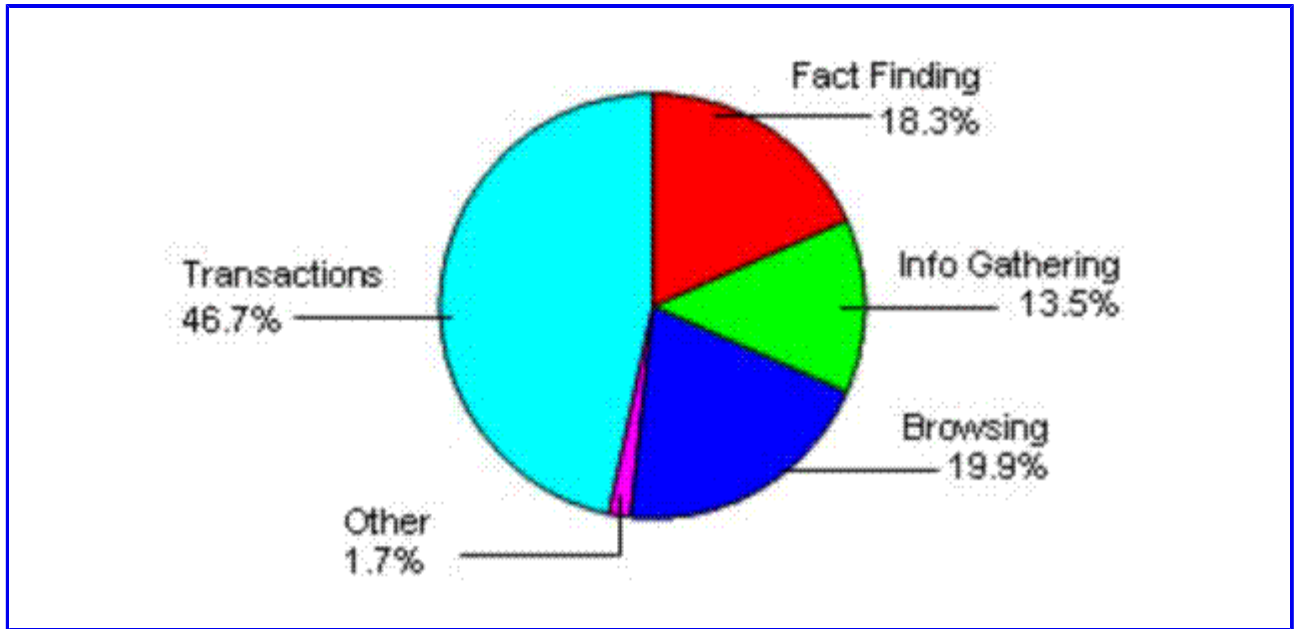


Figure 4. Breakdown of all tasks captured.

Table 3. An overall breakdown of web use, the amount of repeated tasks, and the most commonly initiated tasks.

Fact Finding	Information Gathering
Looking , Searching, or Checking for:	Looking, Researching, Information on:
A book in the library	A new laptop
A course mark	Admissions information
A file (for download)	Beer distribution
A phone number	Breastfeeding
A recipe	Building a computer system
A research paper	Conferences
An Email address	Health Economics
Assignment information	Help with a virus
Bus schedule	iPod prices
Definitions	Job Searching
Directions to a restaurant	Making a resume
Exam dates	Math tools on the Web
Help with a game	New wireless card
How to reference a memo	Palm OS development
How to tie a tie	Papers on policy-based network
Java documentation	Renting a car
Movie times	Risk analysis
Song lyrics	Summer school courses
The average mass of a bullet	
Weather	
Browsing	Transactions

Looking for, Reading:	Checking:
Blogs	Applying for a credit card
Browsing website	Banking
Comics	Blog
Entertainment	Changing profile information
Friend's Homepage	Document delivery request
Gaming forum	Doing an online test
Link received in email	Email
Listening to music	Logging diet and exercise
News	Online accounts
Movie trailers	Online MSN
MP3s	Online shopping
Updates on movie website	Sending a greeting
Wasting time	Taking part in a survey

Fact Finding

Fact Finding tasks accounted for 18.3% (218/1192) of all web usage. Looking for weather information appeared to be the most common Fact Finding task, accounting for 11.5% (25/218) of tasks in this category. Common Fact Finding tasks included looking for course or assignment related material, song lyrics, and specific software. Fact Finding tasks also appeared to be split between personal and school/work-related activities. Table 4 shows a subset of the Fact Finding tasks collected during the study.

Of the 218 Fact Finding tasks, we found that 55.5% (121/218) were repeated at least once. This category had the lowest proportion of repeated tasks. There appears to be three main reasons why Fact Finding tasks were repeated: monitoring, re-finding, and task variants. When monitoring, participants were looking for specific dynamic information, such as the current weather forecast, updated movie times, or the winning lotto numbers. When re-finding, participants were looking to return to a previously found piece of static information, such as bus schedules or exam dates. Task variants occurred when participants were looking for related pieces of specific information, such as looking for programming resources. One example of this was a participant who labelled two tasks "looking for Java documentation" where in one case he was looking for information on hash tables while in another case he was looking for Java documentation on substrings.

When participants described their Fact Finding tasks through the task diary, they often used terms such as "checking" (e.g., checking the weather), "finding" (e.g., finding a phone number), "looking" (e.g., looking for Win XP Pro upgrade), and "searching" (e.g., searching for a journal) for information. The ways in which participants describe their tasks also help to characterize the task itself. These terms and task descriptions reflect the specific

search (finding, searching, looking) or sometimes monitoring (checking) nature of the task.

Information Gathering

Information Gathering tasks accounted for 13.4% (160/1192) of all web usage. There was no single representative task but common tasks included job hunting, course or project related research, researching a new purchase (such as a computer or iPod), and course/admissions information. Many of the Information Gathering tasks were related to technology concepts, which may be attributed to our particular sample population. Table 4 shows a subset of the Information Gathering tasks collected during the study.

For Information Gathering tasks, 58.8% (94/160) of tasks were repeated at least once. Information Gathering tasks appeared to be repeated because participants continued with their tasks at a later time. Since Information Gathering tasks tend to be longer in duration, they were often broken up over a day or even over several days. Among some participants, we saw Information Gathering tasks that stretched over as many as six days, such as a participant who was researching graduate school admission information.

Participants who completed Information Gathering tasks often used terms such as “finding” (e.g., finding information for a presentation), “information” (e.g., information for a project), “looking” (e.g., looking for a new laptop), “research” (e.g., risk analysis research), and “searching” (e.g., searching for iPod prices) to describe their task. While some of the terms used are common to Fact Finding task descriptions (finding, searching), terms such as information and research highlight the collection of information that takes place during information gathering tasks.

Browsing

Browsing tasks accounted for 19.9% (237/1192) of all web usage. Browsing tasks appeared to be related to entertainment or personal interests and consisted of news reading in 40.5% (96/237) of tasks in this category. Other common tasks included reading blogs, visiting gaming related sites, and reading music/TV/movie related web pages. Table 4 shows a subset of the Browsing tasks collected during the study.

Browsing tasks were highly repetitive as 84.4% (200/237) of tasks were repeated at least once. Browsing tasks were primarily habitual or monitoring tasks, such as checking the news or a friend’s blog. We observed many participants who repeated the same Browsing

tasks daily over the course of the study. Often, these were even completed in a habitual sequence. For instance, each morning a participant might read the news on CNN.com, followed by the sports news on ESPN.com, and then read their favourite comic on comics.com.

Participants often used the words “looking” (e.g., looking for a blog update) and “reading” (e.g., reading the news) to describe their Browsing tasks. Often when participants said they were “looking for” something during Browsing, it was in the context of browsing for a hobby or travel related interest.

Transactions

Transactions were the most frequently recorded task, accounting for 46.7% (557/1192) of all web usage. Transactions were primarily made up of web-based email, accounting for 80.4% (448/557) of all Transactions and 37.6% of all web usage. Other types of Transactions recorded by our participants included online bill payments and blog/message board entries. Table 4 shows a subset of the Transactions tasks collected during the study.

We observed that tasks categorized as Transactions appeared to have two distinct goals. The first is the communication of information through email, blog updates, or postings to message boards. The second is the completion of online actions, such as online banking and shopping. This distinction mirrors the task categorization of Communication & Transactions that evolved out of the focus group.

Other

Finally, only a few tasks were categorized as Other and they accounted for 1.7% (20/1192) of all web usage. These were tasks such as viewing web pages during web page development.

DISCUSSION

Classification of Web Information Tasks

Based on the task data collected during the field study, as well as previous work (Ellis, 1989; Choo, Detlor and Turnbull, 2000; Morrison, Pirolli and Card, 2001; Sellen, Murphy and Shaw 2002), we have developed a classification of web information tasks (shown in

Figure 5). The classification consists of three information goals: information seeking, information exchange, and information maintenance. Web information tasks consist of the set of tasks in which users engage on the Web that deal with some aspect of information, from acquisition, consumption, and distribution of information.

Information seeking tasks consist of Fact Finding, Information Gathering, and Browsing. These are tasks in which the user's goal is to change their state of knowledge (Marchionini, 1995). Fact Finding consists of tasks in which the goal is to find a specific piece of information. Information Gathering consists of tasks in which the goal is to collect information, often from multiple sources, in order to write a report, make a decision, or become more informed about a particular topic. Browsing consists of tasks in which there is no specific information goal in mind other than perhaps entertainment or to see what is new.

Information exchange tasks consist of Transactions and Communications. These are tasks in which the user's goal is to exchange information in a web-based setting. Transactions consist of tasks in which an online action takes place, such as banking or a web purchase. Communications consist of tasks that facilitate web-based communication, such as email, online bulletin boards, or web-based publishing such as blog postings.

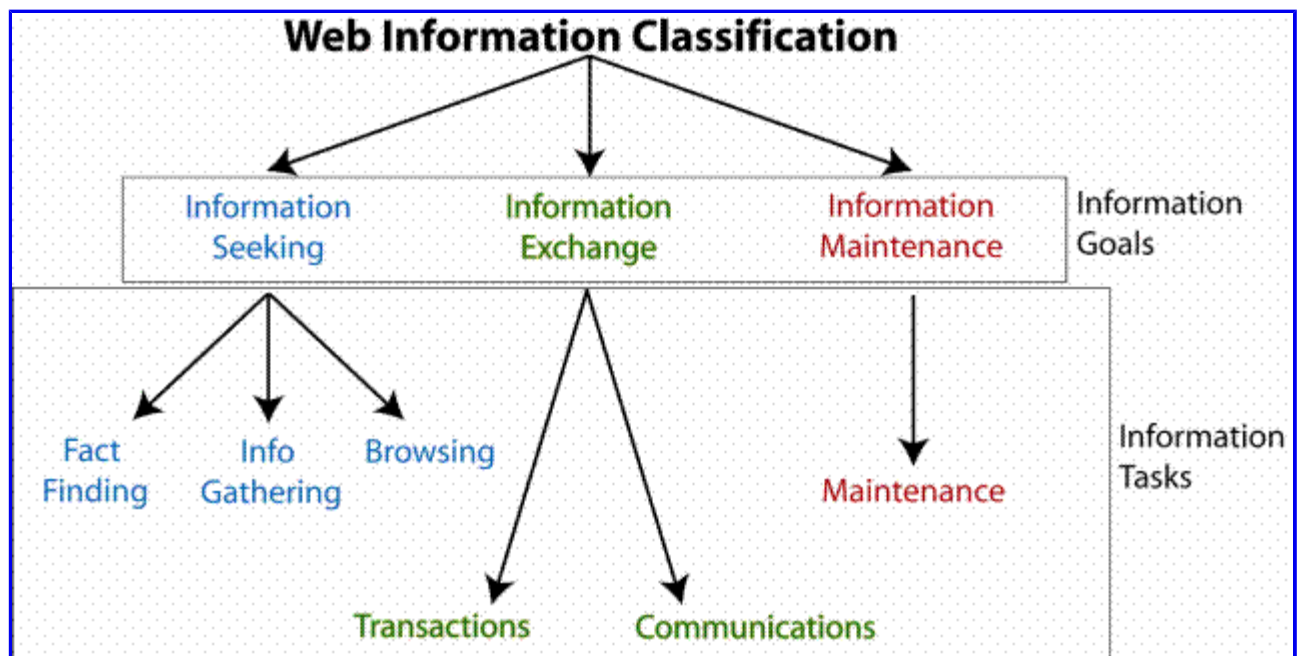


Figure 5. The classification of web information tasks Information maintenance tasks are the tasks which were classified as Other by our participants and as “Housekeeping” by Sellen, Murphy, and Shaw (2002). Maintenance tasks generally consist of visits to web pages with the goal of maintaining web resources, such as to ensure that the content appears as they should, that links are working properly, as well as updates to user profiles.

Most tasks of this nature were observed during the field study when participants were updating or creating new web pages.

This classification has not been built to the exclusion of previous work, but instead has incorporated the components of many previous models and frameworks. In addition, this classification also encompasses other non-information seeking tasks. The collection of six tasks (Fact Finding, Information Gathering, Browsing, Communications, Transactions, and Maintenance) closely mirror the work of Sellen et al. (2002), although our research was conducted with a slightly different focus. Our classification expresses user activities in terms of web information goals and provides a validation of Sellen et al.'s previous framework. Whittaker, Terveen, and Nardi (2000) stated that researchers often tend to conduct a small number of pioneering studies within a task domain, with little or no future follow up by other researchers. Therefore, it is important that researchers continue to validate and iterate on previous studies of user activity on the Web.

Monitoring

One activity that warrants further investigation is Monitoring. We define Monitoring in the same way as previous work (Choo, Detlor and Turnbull, 2000; Morrison, Pirolli and Card, 2001), that is, as an activity that occurs when users return to previously visited pages in order to obtain updated or dynamic information. We were unable to closely study Monitoring because it was difficult to clearly categorize. This study simply gave us an informal view of Monitoring, with no details on what information was being monitored or the goal of the Monitoring activity.

Based on the observations collected during our research, we hypothesize that Monitoring occurs with differing frequency across many web information tasks. We observed high levels of Monitoring within Browsing and Transactions and lower levels within Fact Finding and Information Gathering tasks. However, as previously stated, the study was not designed to study Monitoring and therefore we did not collect a precise set of Monitoring data.

We also expect that the type of information being monitored is dependant upon the higher level web information task. That is, Monitoring within Fact Finding tasks deals with specific, fact oriented information. Information Gathering tasks deal with general, topics based information. Browsing tasks consist of serendipitous, less specific information while Communications consists primarily of email. We also do not anticipate a high level of

Monitoring within Transactions as they consist primarily of online actions and form submissions.

We would like to further study whether Monitoring is in fact an independent information seeking task or simply an activity across all web information tasks. In order to do so, we plan to conduct a series of semi-structured interviews to further investigate users' Monitoring behaviour on the Web.

CONCLUSIONS AND FUTURE WORK

We conducted a field study examining user activities on the Web. Based on the task descriptions and categorizations collected during the study, as well as previous work, we have developed a classification of web information tasks. This classification is based on three main web information goals: information seeking, information exchange, and information maintenance.

The eventual goal of this work is to build more effective tools to support users in their web usage. In order to do so, we must first characterize patterns of users' behaviour on the Web. That is, we need to gain a solid understanding of the types of tasks in which users engage on the Web. Future work will attempt to further understand how monitoring fits into the classification of web information tasks.

ACKNOWLEDGMENTS

This research is supported by Natural Sciences and Engineering Research Council of Canada (NSERC). We would like to thank the members of the WIFL and EDGE research labs. We are also grateful to the 21 participants who took part in the field study.

References

- Belkin, N. J. (1980) Anomalous States of Knowledge as a Basis for Information Retrieval *The Canadian Journal of Information Science* 5: 133-143
- Catledge, L. D. & Pitkow, J. E. (1995) Characterizing Browsing Strategies in the World-Wide Web *Proceedings of the Third International World-Wide Web Conference, Darmstadt, Germany* 1065 - 1073
- Choo, C. W., Detlor, B. & Turnbull, D. (2000) Information Seeking on the Web: An Integrated Model of Browsing and Searching *First Monday* 5(2) http://firstmonday.org/issues/issues5_2/choo/index.html

- Ellis, D. (1989) A Behavioural Approach to Information Retrieval System Design
Journal of Documentation 45(3): 171-212
- Ellis, D., Cox, D. & Hall, K. (1993) A Comparison of the Information Seeking Patterns of Researchers in the Physical and Social Sciences *Journal of Documentation* 49(4): 356-369
- Fenstermacher, K. & Ginsburg, M. (2003) Client-Side Monitoring for Web Mining
Journal of the American Society for Information Science and Technology 54(7): 625-637
- Ford, N., Wilson, T. D., Foster, A., Ellis, D. & Spink, A. (2002) Information Seeking and Mediated Searching. Part 4. Cognitive Styles in Information Seeking *Journal of the American Society for Information Science and Technology* 53(9): 728-735
- Hawkey, K. & Inkpen, K. (2005a) Privacy Gradients: Exploring Ways to Manage Incidental Information During Co-Located Collaboration *Proceedings of the CHI '05 Extended Abstracts on Human Factors in Computing System, Portland, OR* 1431 - 1434
- Hawkey, K. & Inkpen, K. (2005b) Web Browsing Today: The Impact of Changing Contexts on User Activity *Proceedings of the CHI '05 Extended Abstracts on Human Factors in Computing System, Portland, OR* 1443 - 1446
- Herder, E. & Juvina, I. (2004) Discovery of Individual User Navigation Styles
Proceedings of the Workshop on Individual Differences in Adaptive Hypermedia (Adaptive Hypermedia 2004), Eindhoven, The Netherlands
- Hölscher, C. & Strube, G. (2000) Web Search Behavior of Internet Experts and Newbies
Proceedings of the 10th International Conference on World Wide Web, Amsterdam, The Netherlands 337-346
- Kellar, M., Watters, C. & Shepherd, M. (2005) A Field Study Characterizing Web-Based Information Seeking Tasks. *Dalhousie University, Halifax, NS. Technical Report No. CS-2005-20*
- Kellar, M., Watters, C., & Shepherd, M. (2006) The Impact of Task on the Usage of Web Browser Navigation Mechanisms *Proceedings of Graphics Interface 2006, Quebec City, Canada* 235-242
- Kuhlthau, C. (1991) Inside the Search Process: Information Seeking from the User's Perspective *Journal of the American Society for Information Science and Technology* 42(5): 361-371
- Marchionini, G. (1995) *Information Seeking in Electronic Environments* New York, NY, Cambridge University Press

- Meho, L. & Tibbo, H. (2003) Modeling the Information-Seeking Behaviour of Social Scientists: Ellis' Study Revisited *Journal of the American Society for Information Science and Technology* 54(6): 570-587
- Morrison, J. B., Pirolli, P. & Card, S. K. (2001) A Taxonomic Analysis of What World Wide Web Activities Significantly Impact People's Decisions and Actions *Proceedings of the CHI '01 Extended Abstracts on Human Factors in Computing Systems, Seattle, WA* 163-164
- Pitkow, J. E. & Kehoe, C. M. (1996) Emerging Trends in the WWW User Population *Communications of the ACM* 39(6): 106 - 108
- Rieh, S. Y. (2003) Investigating Web Searching Behaviour in Home Environments *Proceedings of the American Society for Information Science and Technology, Long Beach, CA* 255-264
- Rozanski, H. D., Bollman, G. & Lipman, M. (2001) Seize the Occasion! The Seven-Segment System for Online Marketing *Strategy & Competition*
<http://www.strategy-business.com/press/article/19940?pg=19940>
- Sellen, A. J., Murphy, R. & Shaw, K. L. (2002) How Knowledge Workers Use the Web *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Minneapolis, MN* 227-234
- Whittaker, S., Terveen, L. & Nardi, B. A. (2000) Let's Stop Pushing the Envelope and Start Addressing It: A Reference Task Agenda for HCI *Human Computer Interaction* 15: 75-106
- Wilson, T. D. & Walsh, C. (1996) *Information Behaviour: An Interdisciplinary Perspective* Sheffield: University of Sheffield Department of Information Studies