

Chapter 4

Method

In this chapter the method of the study is described including the selection and recruitment of study participants, and the selection of an optimal time period and measurement schedule for the study. The various instruments and materials used to collect data are then described including both electronic and paper instruments, along with specific techniques for measurement. This chapter concludes with a discussion of the procedure used for collecting data.

4.1 Subjects

Seven subjects were recruited to participate in this study. Subjects were Rutgers University graduate students, who were recruited informally through word-of-mouth. Subjects were told that the study was a longitudinal, naturalistic observation of their online information-seeking behaviors and that it would last for a university semester. An attempt was made to select subjects from different Ph.D. programs to allow the subjects to be as heterogeneous as possible in their information-seeking interests and activities.

Subjects represented a variety of Ph.D. programs at Rutgers: history, geography, political science, electrical engineering, communication, mechanical engineering and comparative literature. Subjects were chosen for reasons of convenience, but more importantly, because it was believed that subjects from Ph.D. programs would be likely to engage in a variety of information-seeking activities, which would last for varying amounts of time. Seven subjects were recruited because pilot testing indicated the possibility of a high dropout rate; five subjects were the minimum number of subjects

desired. Further demographic characteristics of the subjects and their experiences in the study are described in Chapter 5.

As participants in the study, each subject received a new laptop computer and printer, which are described in greater detail in Section 4.3.1. Upon completion of the study, subjects were allowed to retain the laptop and printer as compensation for their participation. Subjects who were unable to complete the study were required to return the laptop and printer and issued \$20.00 for each completed week of the study. All subjects completed the study. Subjects were informed during recruitment and at the beginning of the study, both orally and through a Participant Consent Form (see Section 4.3.3), that all of the activities that they performed while using the laptop computer, including online searching, email and word processing, would be logged.

4.2 Time Period

The time period for this study, a single university semester, was selected since it provided a useful temporal division for subjects, who were university students. While not all of subjects' activities were structured around this time period and many activities extended beyond the semester, the university semester provided a convenient time demarcation for this study. This study lasted for fourteen weeks. The study commenced during the week of January 27th 2003 and ended during the week of May 12th 2003.

4.3 Instruments and Materials

This section describes the study laptops, the various pieces of logging software, the consent form, the content of the various questionnaires administered during the study,

the software used by subjects to evaluate pages and capture these evaluations, and the protocol of the Exit Interview. All instruments and materials described in this section, with the exception of the laptops and printers, were piloted on numerous occasions, both formally and informally, with participants other than those of this study. Prior to distribution to the subjects, the laptops were only used by the investigator to install and test software.

4.3.1 Laptops and Printers

Each subject received a new Gateway Solo 1450 laptop equipped with Windows XP, 256 MB RAM, 30 GB of hard disk, and a 1.33 GHz Celeron processor. Each laptop had an internal modem/LAN card, DVD/CD-ROM and floppy disk drives. All laptops were equipped with the standard software provided by Gateway, including Internet Explorer and Norton Anti-virus software. In addition, the laptops were equipped with the Microsoft Office Suite (Microsoft Word, Microsoft Excel, Microsoft PowerPoint) and Microsoft FrontPage. The laptops were also equipped with WSFTP, a program for transferring files, Adobe Acrobat, Netscape Communicator and various pieces of logging software, which will be described in the next section.

Subjects were allowed to install other software on the laptop with permission from the investigator. Assistance was offered to subjects with establishing a remote dial-up connection, although most subjects did not require assistance. Subjects were provided with all of the materials that came with the laptop except for the bill of sale, which was retained by the investigator until subjects completed the study. Subjects were also provided with a new Epson Stylus printer.

4.3.2 Logging Software

The laptops were equipped with the WinWhatWhere Investigator client-side logging software. Subjects' online activities were also directed through a proxy logger. Each application is described in more detail below.

WinWhatWhere Investigator (www.winwhatwhere.com) is a commercially available piece of monitoring software whose primary clientele are those who are interested in secretly spying on the online activities of other people, such as employees, spouses or children. Given the costs associated with building a custom logger, both in terms of time and money, it was determined that a commercial product was the most tractable option for logging. WinWhatWhere Investigator was chosen after a thorough review and evaluation of various other pieces of commercially available loggers.

WinWhatWhere Investigator was launched automatically each time the subject's laptop was started and executed in stealth mode while the laptop was in operation. The software did not interfere with any of the subject's natural behaviors; instead, the software unobtrusively monitored and recorded subjects' interactions with all applications including the operating system, web browsers and word processors. Information such as applications used, URLs visited, start, finish and elapsed times for interactions and all keystrokes, including queries, were recorded. This output consisted of:

- Date – the date that the action occurred. Example value: 4/23/2003.
- Start – the time that the action started. Example value: 10:39:24 AM.
- Elapsed – the difference in time between the start time of the current action and the start time of the preceding action. Example value: 00:00:35.

- Exe – the location of the application used to execute the action. Example value: C:\Program Files\Internet Explorer\iexplore.exe.
- Caption – the caption of the item being executed.
 - For Web pages viewed, this included the URL of the page and/or the title given to the page by the web page creator. Example values: Weightloss and Fad Diets - the Cabbage Soup Diet and <http://weightloss-and-diet-facts.com/dietmyths.htm>
 - For word processing documents, this included the file name given to the document by the user. Example value: weightloss.doc.
 - For operating system (OS) activities, this included the caption given by the OS. Example value: CREATE:Weightloss and Fad Diets - the Cabbage Soup Diet.
- Keystrokes – the number of keystrokes made on a particular document. Example value: 9.
- Formatted – the textual (letters and numbers) keystrokes made at a particular document. This excluded keystrokes made for scrolling, moving the cursor, and pressing return. For queries, this was the final query submitted upon pressing enter, excluding any corrections or deletions made before the query was submitted. Example value: dowco motorcycle cover.
- Raw - the exact sequence of keystrokes made at a document. This included any corrections or deletions, use of keys for scrolling, moving the cursor backwards and forwards and pressing return. Example value: m<BACK>dowco
otr<BACK><BACK><BACK>motorcycle cover

Display time and retention were collected from the WinWhatWhere logger. Display time was the length of time that a document was displayed in the subject's active web browser window. Elapsed time was used to measure display time for documents. The logger recorded this as the difference in start times from successive entries. For example, if a subject went to the Google home page (D1), entered a query and pressed search, reviewed the search results (D2) and then followed one of the links (D3), the elapsed time for D1 would be the difference in start times from D1 and D2. The elapsed time for D2 would be the difference in start times from D2 and D3. For identical pages viewed at different times (for instance, if the user above went back to the search results page), elapsed time entries were summed to arrive at the total elapsed time. These computations were made on a weekly basis, corresponding to each week of the study. Since the client-side logger recorded all of the subject's interactions including those with applications other than a web browser, it was possible to determine the correct elapsed time for web documents even when the subject was multi-tasking between a web browser and another application such as a word processor. Also, if more than one web browser was being used or if more than one browser window was opened, the logger identified all separate instances of the web browser.

Retention behaviors of interest in this study were printing, saving, and bookmarking. When a subject performed one of these actions, the logger recorded this action as a new entry underneath the document entry which received the action. Printing, saving and bookmarking were each recorded as separate variables and coded in binary form, where "1" represented the presence of the action and "0" represented the absence of the action.

The proxy logger was a custom built logging application that resided on a local proxy server that was used to save a local copy of each page request made by subjects. Subjects' Web browsers were directed through the proxy logger; this direction did not disrupt subjects' activities or cause any noticeable lag times. The proxy logger was developed at the University of Massachusetts, Amherst as part of the Mongrel Project with Rutgers University (www.scils.rutgers.edu/mongrel).

Each time a subject initially requested a page through the proxy, the proxy would either create a new folder in which to store the request, and subsequently requested documents, based on the IP address of the subject or it would add the page to a pre-existing folder. If the subject had a permanent connection to the Internet, such as a DSL or LAN connection, the subject had a folder named after this static IP connection. For each page requested during a particular session, the proxy would add the newly requested page to a .trec file contained within the IP-named folder. All pages requested during the connection period were appended to this file. If the subject disconnected and then reconnected at a later time during the day, then a new folder would be created based on the new IP address, and a new .trec file would be created. The .trec files were named according to the current date.

The .trec file format is the format used for web documents by the TREC web track (<http://trec.nist.gov>). This file format appends requested documents with several pieces of metadata. Figure 4.1 presents an example of the metadata tags that were added to each document by the proxy logger. Because all documents requested during a particular session by a particular subject were appended to the same .trec file, a <DOC> </DOC> tag set surrounded the document, which facilitated the subsequent parsing of the

individual documents. Within this tag set, there was a tag set for assigning a unique identifier to the document (<DOCNO>); recording the date, time and URL of the request (<DATE>, <TIME>, <URL>); indicating from which, if any, previous document the subject clicked to display the current document (<REFERER>); and indicating the IP address from which the request came (<USERID>). The normal, raw marked-up text of the requested document followed these tag sets. This included both html and non-html documents, such as cascading style sheets.

```

<DOC>
<DOCNO>HZ128.6.251.4303052003150931-91</DOCNO>
<DATE>03052003</DATE>
<TIME>150931</TIME>
<URL>http://www.sloan.org/programs/index.shtml</URL>
<REFERER>http://www.sloan.org/main.shtml </REFERER>
<USERID>128.6.251.43</USERID>

<HTML>
<TITLE>Alfred P. Sloan Foundation</TITLE>
<BODY>
.
.
.
</HTML>
</DOC>

```

Figure 4.1 Output of the proxy logger (.trec file format)

For those subjects who dialed-up to establish a connection to the Internet or used varying LAN connections provided around campus, there was no way for the investigator to know which IP address belonged to which subject since this information was variable. To establish unique identities for each subject within the .trec files, unique html pages were created for each subject. These pages were set as the subjects' browser home page and subjects were asked to maintain this page as the home page throughout the study. Subjects were also asked to make a "hard" reload of the page each time that they

connected. A “hard” reload ignores any locally cached copy of the requested page and issues a new request for the page. This insured that requests for the page passed through the proxy and were logged in the .trec file. For most cases, this method worked well in identifying which .trec files belonged to which subject.

4.3.3 Consent Form

Before receiving the laptop and printer, each subject was required to read and sign an informed consent form (Appendix A). The consent form described the: (1) general purpose of the study; (2) length of the study; (3) distribution and use of the laptop and printer; (4) all laptop activities that would be monitored and recorded; (5) the general conduct of the study, including weekly meetings; and, finally, (6) compensation for participation. The consent form further described to the subject how the collected data would be analyzed, stored and used in the future. Each subject received a signed copy of the consent form for their personal records. Approval from the Rutgers Institutional Review Board (IRB) for the Protection of Human Subjects was obtained for this research project and approval was granted prior to the start of the study. This approval is documented under the IRB protocol titled Information Search and Use Behaviors as Implicit Sources of User Preference and States of Knowledge, Protocol #02-269M.

4.3.4 Laptop and Printer Agreement

Subjects were asked to sign a Laptop and Printer Agreement (Appendix B) that detailed their responsibilities as caretakers of the laptop and printer, and stipulated the conditions under which the equipment should be returned to the investigator.

4.3.5 Entry Questionnaire

The Entry Questionnaire (Appendix C) gathered background and demographic information from subjects and questioned subjects about their previous computer and searching experiences. The information obtained from the Entry Questionnaire was used to characterize the subjects, but not in subsequent data analysis. Subjects were asked to indicate their academic majors and degrees earned, Ph.D. program currently enrolled, gender, and age. Subjects were also asked to provide a short statement of their general interests, including academic, career and recreational interests.

4.3.6 Task and Topic Questionnaires

The Task and Topic Questionnaires (Appendix D and E) elicited the tasks and topics that were of current interest, or were expected to be of interest, to the subject during the study. Subjects were asked to think of their online activities in terms of tasks and topics. Example tasks and topics were provided to subjects,

“An example task might be writing a research paper. The topic of this task might be information retrieval and/or interfaces. Another example task might be travel. The topic might be Oregon or Paris. Another task might be shopping, with the topic being shoes or clothes.”

Subjects were told that a single task might have numerous topics or that a single topic might be related to numerous tasks. Subjects were instructed that there were no correct tasks and topics and that their identification of tasks and topics was completely subjective. Subjects were further instructed that some activities might be difficult to classify according to task and topic, but that they should try their best. While it was not

always easy for subjects to make these distinctions, results from several pilot tests demonstrated that subjects could do this consistently, with few problems.

The Task and Topic Questionnaires asked subjects to characterize each task and topic that was identified according to several attributes. For each task, subjects were asked to indicate endurance, frequency, and stage. For each topic, subjects were asked to indicate persistence and familiarity. The Task and Topic Questionnaires were completed at the initial meeting with the investigator, following the completion of the Entry Questionnaire. These attributes, along with a description of how they were measured, are presented below.

- Task endurance was the length of time the subject expected to be working on the task. Task endurance was measured on an eight-point scale, whose eight points demarcated specific lengths of time. The determination of an eight-point scale was based on the time periods of interest, and by empirical work (Tang, Shaw, Vevea, 1999) that found that seven-point scales provide an optimal range of responses to subjects. Because the time demarcations corresponded to conventional measures of time (day, week, month and year), eight points were used instead of seven. These eight points were (with their corresponding numeric values in parenthesis): one day (1); several days (2); one week (3); several weeks (4); one month (5); several months (6); one year (7); and several years (8).
- Frequency was how often the subject expected to conduct online information-seeking activities related to a task. As with task endurance, frequency was measured on a eight point scale, whose eight points demarcated specific amounts of time (with corresponding numeric values in parenthesis): once or twice a day

(1); daily (2); once or twice a week (3); weekly (4); once or twice a month (5); monthly (6); once or twice a year (7); yearly (8). Several subjects noted during the study that differences between points 1 and 2; 3 and 4; 5 and 6; and 7 and 8 were not obvious; indeed, these points do not seem to be orthogonal. Thus, for analysis, these pairs of eight points were collapsed into four.

- Stage was subjects' assessment of their progress in completing the task. Stage was measured on a seven-point scale based on the study cited earlier (Tang, Shaw, & Vevea, 1999) that found that seven point scales provide an optimal range of responses to subjects. It was also desirable to have an odd-numbered scale, so that a mid-point was possible for this variable. The scale had labels for the two anchors, starting (1) and finished (7), and numeric values for intervening values. Because the stage model of information-seeking (Kuhlthau, 1992) is based on a specific type of task (writing a research paper), subjects were provided with the option of "not applicable" for this scale since not all of their tasks were of this type or had a specific stage.
- Persistence was the length of time the subject expected to be interested in information about a topic. Persistence was measured on an eight-point scale, whose eight points demarcated specific lengths of time. These eight points were (with their corresponding numeric values in parenthesis): one day (1); several days (2); one week (3); several weeks (4); one month (5); several months (6); one year (7); and several years (8). As with task endurance, the determination of an eight-point scale was based on the work by Tang, Shaw and Vevea (1999) and conventional measures of time.

- Familiarity was the subject's current state of knowledge about a topic. Topic familiarity was measured on a seven-point scale that had labels for the two anchors, unfamiliar (1) and familiar (7), and numeric values for intervening values. The choice of a seven-point scale was guided by the work of Tang, Shaw, and Vevea (1999) and the desire to have a mid-point on the scale.

4.3.7 Task and Topic Update Questionnaires

The Task and Topic Update Questionnaires (Appendix F and G) were similar to the Task and Topic Questionnaires described in the preceding section, with some minor exceptions. These questionnaires were administered at weekly intervals following the initial meeting. The Task and Topic Update Questionnaires were customized to each subject in that they presented the subject with all of their previously identified tasks and topics and their characterizations of each of these tasks and topics according to endurance, frequency, stage, persistence and familiarity. Subjects were asked to update the lists by eliminating tasks with which they were no longer working and topics in which they were no longer interested, and re-characterizing all other tasks and topics according to the attributes. Subjects completed Task and Topic Update Questionnaires each week of the study. The process of adding new tasks and topics is discussed below.

Initially, Task and Topic Update Questionnaires were designed such that all tasks and topics were presented on unique pages along with the previous week's characterizations. In managing the creation of these questionnaires each week for each subject, numerous problems were encountered, some of which resulted in minor errors. It was also observed that an unnecessarily large amount of paper was being consumed and

that the handling of the stacks of paper by subjects was cumbersome and awkward. Based on these observations, the format of these questionnaires was changed during Week 7 of the study (see Appendix H and I). The new format still displayed the previously identified tasks and topics to subjects along with their characterizations, but this information was displayed in tabular format. Each task and topic, along with the characterizations from all previous weeks, was displayed in a single row. Subjects were happy with this change and enjoyed being able to see the history of the tasks and topics and their respective ratings. This change greatly decreased the amount of work required of the investigator. It is believed that this change in format did not adversely affect subjects' task and topic identification and characterization activities.

When adding a new task or topic, subjects were asked to complete a Task or Topic Questionnaire online, which was identical to those described in the previous Section (Appendix D and E). The addition of new tasks or topics occurred after the subject completed Update Questionnaires. The reason for doing the additions online was so that each new task and topic could be automatically added to a personalized list of tasks and topics that the subject used during their page evaluation activities (described in the next section). Page evaluation activities occurred immediately following Task and Topic Update Questionnaires; completing these questionnaires online allowed new tasks and topics to be automatically added to the list. Alternatively, subjects could add a new task or topic during their page evaluation activities using the same version of the electronic questionnaires described above. In most cases, subjects added new tasks and topics during their page evaluation activities.

4.3.8 Evaluation Software

For each week of the study, after completing the Task and Topic Update Questionnaires, subjects were presented with a selection of the documents that they had requested during the previous week and were asked to: (1) classify each document according to their tasks and topics; (2) indicate the usefulness of the document as it related to that task and topic; (3) indicate their confidence in the usefulness rating that they assigned to the document. If subjects could not remember a document, they were instructed not to evaluate the document.

Usefulness was measured on a seven-point scale where the scale anchors were “not useful” and “useful.” Numeric values were not provided on the scale, so that the scale points appeared more continuous, rather than discrete, to subjects. Data for each point was coded for analysis with numeric values from 1 to 7, where “1” indicated “not useful” and “7” indicated “useful.” Subjects could also indicate that they were “unsure” of the usefulness of the document. As with stage and familiarity, the determination of a seven-point scale for usefulness was based on the work of Tang, Shaw, & Vevea (1999) who recommend the use of scales with six to seven points for relevance evaluation. Based on the work of Tang, Shaw & Vevea (1997), it was concluded that a seven-point scale would insure that participants had a representative range with which to evaluate documents and that the reliability of the scale would be maximized.

Confidence was the extent to which the subject believed that the usefulness rating that they assigned to a document reflected their opinion of the document’s usefulness. Confidence was measured on a seven-point scale, where the scale anchors were “low ”

and “high.” Responses were coded with numeric values from 1 to 7, with “1” representing “low” and “7” representing “high.”

Evaluations were conducted using evaluation software that was developed in-house at Rutgers University as part of the Mongrel Project. The evaluations took place on a Dell Inspiron 3800 laptop. The interface for this software is displayed in Figure 4.2.

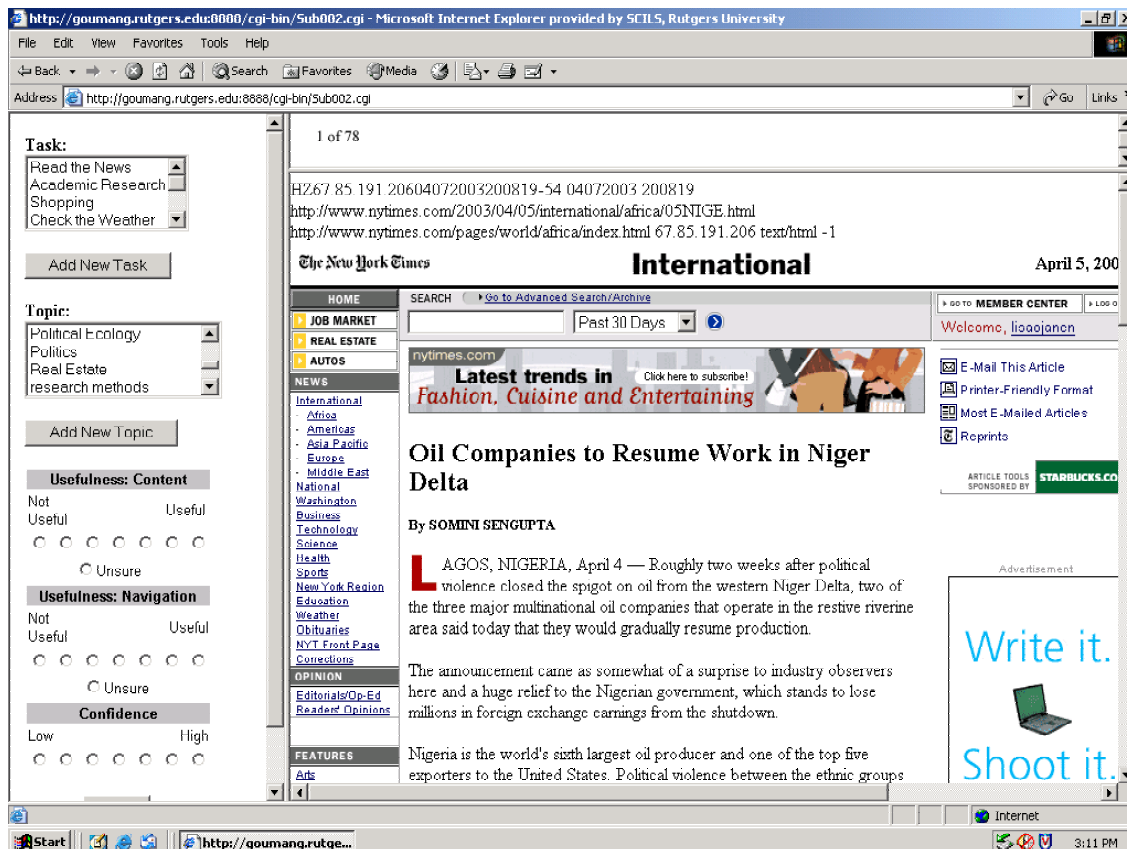


Figure 4.2 Evaluation interface

The interface presented subjects with the following features:

- Two personalized lists of the subject’s tasks and topics in separate, scrollable windows. To select a task or topic, the subject clicked on the task or topic. Two or more tasks or topics could be selected by holding down the control key while clicking on the tasks or topics.

- Buttons that allowed the subject to add a new task or topic to the list. Clicking on these buttons opened a separate browser window which displayed a Task or Topic Questionnaire. Upon completing and submitting this questionnaire, the newly added task or topic was appended to the pre-existing list of tasks or topics.
- A scale for the subject to indicate the usefulness of the document with respect to the document's content. This scale was implemented as a series of seven radio buttons. Subjects were also given the option "unsure."
- A scale for the subject to indicate the usefulness of the document with respect to its navigation. This scale was also implemented as a series of seven radio buttons. Again, subjects were provided with the option of "unsure."
- A scale for the subject to indicate their confidence in the usefulness judgment. Like the two usefulness scales, the confidence scale was also implemented as a series of seven radio buttons.
- A small frame displaying the number of evaluations made during the session and the total number of documents that remained to be evaluated.
- A frame displaying the content of the document.

The documents shown to subjects during the evaluations were stored on, and served from, a local machine. For practical and theoretical reasons, not all of the documents requested by subjects were shown to subjects for evaluation. Indeed, individual document requests exceeded one-thousand during some weeks. Several techniques were established for selecting which documents would be displayed to subjects. The first technique was a content-based classification scheme of page types that was created based on a manual examination of 2,000 web documents collected during a

pilot study (Appendix J). The development of this classification scheme was based on the work of Haas and Grams (2000) who constructed a classification of web page types by performing a content analysis of a number of web pages. The purpose of the classification scheme as used in this study was to identify and select systematically the documents that would be evaluated by subjects. Two people, who were not participants of the current study, validated this classification scheme through inter-coder reliability tests.

All of the documents requested by subjects were evaluated according to this classification scheme. The development of this classification scheme was based on the observation that not all documents are equally useful for establishing the subject's interests, and that document structure can often influence the kinds of behavior one is likely to exhibit at a document. For example, many subjects viewed the Google homepage. Unless customized, this page is identical for every subject and there is no content on the page that could contribute to the understanding of the subject's interests. Furthermore, it makes little theoretical sense to ask the subject to classify the Google home page according to task or topic. While queries certainly have the potential for contributing to an understanding of the subject's interests, they were not of concern in this study. Other document types were excluded because their structure could influence the subject's behavior. For instance, one might expect longer display times, a greater number of click-throughs and more frequent displays for search results pages or a discussion group index page. Moreover, search results pages contain a list of pointers, some more relevant than others, and many not relevant at all. Evaluating the usefulness of these types of pages again makes little theoretical sense.

Email pages, chat room discussions, news groups, bulletin boards, online games and pornography were not selected as document types to show to subjects to provide them with some privacy and to narrow the scope of this study. Ultimately, two basic kinds of documents were shown to subjects: (1) homepages that consisted of more than just search facilities and (2) documents which did not fit into any other category. For example, the Google homepage would not be shown, but the SCILS homepage would be shown. For (2), one can image a search record, such as those found when searching the Rutgers Library Online Catalog or a specific news article.

Several other considerations were made after the documents had been screened according to the classification scheme. Documents of all types not written in the English language were excluded because ultimately, to make use of the content of the page, one would need to employ cross-language IR and this was well beyond the scope of this dissertation. Subjects were shown documents that they viewed on multiple occasions during the week only once if the content of that document did not change. If the content did change, the document was considered as a different instance and was shown multiple times. An example of such a document is a current news document, such as the world news page of CNN, whose content usually changes from day-to-day, and sometimes even changes during the same day. In addition, some documents from specific sites did not display from a local location or generated errors for reasons unknown to the investigator. These documents were excluded from the evaluation.

For the first few weeks of the study, all remaining documents after the screening process was completed were presented to subjects for evaluation. However, there was a limit to how many documents subjects could evaluate during the one-hour meeting.

Initially, subjects were instructed to think-aloud during their evaluations (to be described in detail in 4.3.9), but since this process required extra cognitive resources, the rate with which subjects' were able to evaluate documents was reduced. Because of the goals of this dissertation, the think-aloud protocol was eliminated and subjects were able to evaluate almost three times as many documents as they had been able to with the protocol.

If a subject was unable to evaluate all of the documents from one week, those documents were rolled-over to the end of the proceeding week's evaluation session. In a couple of cases, a week where many documents were requested was followed by a week where few documents were requested and this did not present any major problems. However, this was impractical for consecutive weeks where subjects viewed several hundred documents. In these cases, one hundred and fifty documents were randomly selected to be shown to the subject. Subjects were asked during the session to evaluate as many as possible.

4.3.9 Screen Camera

Camtasia (www.camtasia.com) is a PC screen capture application which captures both the contents of the computer screen as well as audio. The result of the capture is saved as a .avi file. Audio was captured by a microphone, which was placed next to the computer. During the weekly evaluations, Camtasia was used to capture subjects' evaluation activities. During the first couple of meetings, subjects were instructed to think-aloud during the document evaluations. That is, subjects were asked to verbalize what they were thinking during the evaluations, and to specifically indicate why they

were classifying and rating documents in a particular way. However, it was quickly discovered that subjects who were good at thinking-aloud (i.e. those subjects who were vocal), were unable to evaluate many documents (approximately 30), while those subjects who were not good at thinking-aloud (i.e. those subjects who were not vocal) were able to evaluate almost three times as many documents during the same period of time. Because of the goals of this study, it was decided that subjects would not be instructed to think-aloud to enable them to evaluate a greater number of documents. However, Camtasia continued to be used during the evaluations to capture the contents of the screen and any comments that subjects might contribute.

At the last document evaluation session, subjects were presented with fewer documents and asked to think-aloud during their evaluations to gain some insight into their process of classifying and rating. Subjects were asked to indicate why they were classifying and rating documents in a certain way and how confident they were about their evaluations.

4.3.10 Exit Interview

The Exit Interview was designed to be an open-ended discussion of subjects' tasks, topics and characterizations of each of these, as well as subjects' perceptions and interpretations of the scales and procedures used during the study. The purpose of the Exit Interview was to: (1) obtain more detailed descriptions of each task and topic by asking subjects to describe each one in turn, providing specific examples; (2) have subjects orally reflect on the activity of classifying documents into task and topic classes; (3) have subjects orally reflect on the history of their task and topic characterizations and

explain these histories; and (4) obtain comments on subjects' interpretations of the usefulness and attribute scales. The Exit Interview schedule can be found in Appendix K. All Exit Interviews were audiotaped with a micro-cassette recorder and transcribed for analysis.

4.4 Procedure

At the first meeting in the office of the investigator, subjects were told that they were participating in a study whose goal was to understand how online information-seeking behavior could be used as implicit evidence of interests and document preferences. Subjects were told that they would receive a new laptop and printer to use during their participation in the study and would be allowed to keep them as compensation, once they completed the study. Subjects were asked to read and sign a consent form that described the activities that they would be asked to engage in throughout the study. The form also informed them of the monitoring software installed on the laptop and proxy, and explained their rights as subjects. After signing the consent form, subjects read and signed the Laptop and Printer Agreement. Subjects then completed the Entry Questionnaire, which gathered demographic and background data, and the Task and Topic Questionnaires, which elicited their current tasks and topics and characterizations of each of these.

Once a subject completed these activities, a laptop and printer were distributed to the subject. The laptop was turned on and demonstrated to subjects by the investigator. Descriptions of the various applications and functionalities were given to subjects. Subjects were given all materials that came with the laptop, except for the bill of sale,

which was retained by the investigator until the subject completed the study. Finally, the weekly evaluation sessions were scheduled either at the close of first session or via email on the day following the initial meeting. Subjects were instructed to bring their laptop to each weekly evaluation session.

For each week of the study, subjects were required to participate in a one-hour meeting where they completed the Task and Topic Update Questionnaires and evaluated the documents that they viewed during the previous week. It was decided to conduct these activities on a weekly basis since scheduling them at more frequent intervals would be too burdensome for both the subject and the investigator, and scheduling them at less frequent intervals would prohibit the subject from evaluating many of the pages that they viewed because of the quantity of documents to be considered. Having subjects evaluate documents as they viewed them was considered, but it was deemed to be too disruptive. Results from a pilot test indicated that weekly intervals provided the most effective and efficient time interval for evaluations.

During the first week of the study, problems were experienced with the proxy logger and logging did not occur. However, weekly evaluation sessions were still held to demonstrate the protocol of the weekly evaluation sessions and the evaluation software. It was decided that Task and Topic Questionnaires from the initial meeting would be treated as a pilot and that the first tasks and topics recorded in the data would be those from the first weekly evaluation session. This information came from Task and Topic Update Questionnaires, which subjects completed at the start of the first weekly evaluation session. Subjects made few changes to these Update Questionnaires. In other words, the information from these Questionnaires was very similar to that provided in the

initial Task and Topic Questionnaires. While subjects completed these Questionnaires, the investigator checked the laptop to make sure that all logging software was functioning properly. After subjects completed the Questionnaires, they were instructed in the protocol for evaluating pages and presented with the evaluation interface, which contained their personalized list of tasks and topics and a set of sample documents collected by the investigator. Subjects were asked to pretend that the documents were those that they had requested during the previous week and to think-aloud while evaluating the documents. The purpose of this session was to familiarize subjects with the evaluation session and protocol. This session was recorded using Camtasia.

Over the course of the next fourteen weeks, subjects returned to the investigator's office once a week for one hour to complete the Task and Topic Update Questionnaires and the weekly document evaluations. During the week of Spring Break, no weekly meetings were held. Instead, subjects were asked to spend an extra thirty minutes evaluating documents at the first meeting following the Break. On occasion, at special request, the dates and times of weekly meetings were changed to accommodate the schedule of the subjects or the investigator.

During each weekly meeting, the laptops were checked to insure that all client-side logging software was functioning properly. The data generated by this software remained on the laptops until the conclusion of the study with one exception. At several points during the study, it was discovered that the WinWhatWhere Investigator was randomly turning itself off for reasons that still remain unknown to the investigator. This caused a loss of several days of data for some subjects. The first solution to this problem was to have WinWhatWhere secretly email the log records to the investigator every

couple of hours, which removed the records from the subject's machine. This process allowed for the identification of long periods of inactivity on the subject's machine. On these occasions, the investigator emailed the subject inquiring about recent activity. If the laptop was in use, but no log records were being emailed, then the subject returned with the laptop to have the investigator reactivate the software.

After two weeks of this, it became apparent that this process was impractical and burdensome for the subject. At this point, a free online windows scheduler (Splinterware <http://www.splinterware.com>) was downloaded to each laptop. The investigator scheduled the start window for WinWhatWhere to pop-up every two hours, where the subject could see if WinWhatWhere was no longer running. When this happened, subjects were asked to reactivate the application by clicking on a Start button found on the WinWhatWhere application window.

The last weekly meeting with subjects (Week 14) was scheduled for two hours. At the start of the meeting, subjects completed Task and Topic Update Questionnaires as they normally would. Subjects then completed their document evaluations. Subjects were presented with fifty or fewer documents and instructed to think-aloud during their evaluations, describing all of the evaluations that they made to these documents. Due to time constraints, subjects were only allowed to evaluate documents for thirty minutes. For one subject, no documents were requested during the previous week. This subject was shown documents from weeks prior to gain some insight into her evaluation process. While the subject evaluated documents, the investigator removed all data and experimental applications from their laptop.

After subjects completed the evaluation, they were asked to comment on the usefulness and confidence scales. Specifically, subjects were asked to identify any limitations of the scales and difficulties that they had using the scales to evaluate documents. Subjects were then presented with their recently completed Task and Topic Update Questionnaire and asked to consider each task and topic one-by-one, and describe, in detail, each one, providing examples when appropriate. Subjects were instructed to identify examples that were problematic or ambiguous. Subjects were further asked when considering each task and topic, to reflect on the history of the characterizations, indicating why something had or had not changed over time. Next, subjects were asked to comment on any difficulties they experienced with thinking of their online searching activities in terms of tasks and topics, creating labels for tasks and topics, distinguishing between tasks and topics, and classifying documents into task and topic classes. Finally, subjects were asked to comment on the scales used to characterize each task and topic. Specifically, subjects were asked to describe their interpretation of the scales and to indicate any difficulties they experienced when using the scales to characterize their tasks and topics.

At the close of the meeting, subjects were debriefed about the study and given the bill of sale for their laptop. Contact information was gathered from each subject and the investigator used this information to transfer the licenses for the computers to each subject through Gateway. This contact information was then destroyed.