ED 438 799   IR 019 927

AUTHOR   Choo, Chun Wei; Detlor, Brian; Turnbull, Don
TITLE   A Behavioral Model of Information Seeking on the
                Web--Preliminary Results of a Study of How Managers and IT
                Specialists Use the Web.
PUB DATE   1998-10-00
NOTE   16p.; Paper presented at the Annual Meeting of the American
                Society for Information Science (ASIS) (61st, Pittsburgh,
                PA, October 25-29, 1998).
PUB TYPE   Reports - Research (143) -- Speeches/Meeting Papers (150)
EDRS PRICE   MF01/PC01 Plus Postage.
DESCRIPTORS   Behavior; Foreign Countries; *Information Scientists;
                *Information Seeking; Internet; Media Specialists; Models;
                *Online Searching; *Search Strategies; User Needs
                (Information); *World Wide Web
IDENTIFIERS   Browsing

ABSTRACT
                This paper develops a new behavioral model of information
seeking on the Web by combining theoretical elements from information science
and organization science. The model was tested during the first phase of a
study of how managers and information technology (IT) specialists use the Web
to seek external information as part of their daily work. Participants
answered a questionnaire and were interviewed individually in order to
understand their information needs and information seeking preferences. A
custom-developed tracker application was installed on their workplace
computers, or their browsers were redirected through a proxy server set up by
the research team. Participants' Web-use activities were then monitored
continuously for two work weeks. The tracker application recorded
participants' Web browser actions, while the proxy recorded HTTP requests and
transfers. In a follow-up round of personal interviews, participants recalled
critical incidents of using information from the Web. Data from the
questionnaire, interviews, and the tracker and server log files supplied a
rich database for study. Results were found to be compatible with the
behavioral model proposed. Overall, the study suggests that a behavioral
framework which relates motivations (strategies and modes of viewing and
searching) and moves (tactics used to find and use information) may be
helpful in analyzing Web-based information seeking. Findings also suggest
that multiple, complementary methods of collecting qualitative and
quantitative data may be used within a single study to compose a richer
portrayal of how individuals seek and use Web-based information in their
natural work settings. (Contains 18 references.) (Author/AEF)

A Behavioral Model of Information Seeking on the Web-Preliminary Results of a Study of How Managers and IT Specialists Use the Web

By Chun Wei Choo, Brian Detlor, Don Turnbull

1998 ASIS Annual Meeting Contributed Paper

A Behavioral Model of Information Seeking on the Web --
Preliminary Results of a Study of How Managers and IT Specialists Use the Web

Chun Wei Choo, Brian Detlor, Don Turnbull

(choo@fis.utoronto.ca, detlor@fis.utoronto.ca, turnbull@fis.utoronto.ca)

Faculty of Information Studies, University of Toronto, Toronto, Ontario, Canada

Abstract

The paper develops a new behavioral model of information seeking on the Web by combining theoretical elements from information science and organization science. The model was tested, in a preliminary way, during the first phase of a study of how managers and IT specialists use the Web to seek external information as part of their daily work. Participants answered a questionnaire and were interviewed individually in order to understand their information needs and information seeking preferences. A custom-developed tracker application was installed on their workplace computers, or their browsers were redirected through a proxy server set up by the research team. Participants' Web-use activities were then monitored continuously for two work weeks. The tracker application recorded participants' Web browser actions, while the proxy recorded HTTP requests and transfers. In a follow-up round of personal interviews, participants recalled critical incidents of using information from the Web. Data from the questionnaire, interviews, and the tracker and server log files supplied a rich database for study. Thirty significant episodes of information seeking were isolated and analyzed in terms of their modes of viewing or searching, and their associated Web information moves. Results were found to be compatible with the behavioral model proposed. Overall, the study suggests that a behavioral framework which relates motivations (the strategies and modes of viewing and searching) and moves (the tactics used to find and use information) may be helpful in analysing Web-based information seeking. The study also suggests that multiple, complementary methods of collecting qualitative and quantitative data may be used within a single study to compose a richer portrayal of how individuals seek and use Web-based information in their natural work settings.

1 Research Objectives

The research presented in this paper has three objectives:

1. To develop a new behavioral model of information seeking on the Web based on a synthesis of theoretical elements from information science and organization science;
2. To test, in a preliminary way, the viability of the model using a modest set of field-data from a pilot study;
3. To experiment with the use of multiple, complementary methods of collecting qualitative and quantitative data on how individuals seek and use Web-based information in their natural work settings.

The paper is organized into five sections. Section 2 outlines two recent conceptual models of information seeking on the Web. Section 3 reviews and combines elements from research in organizational scanning and information seeking into a new behavioral model of Web-based information seeking. Section 4 presents preliminary results from our pilot study which appear to be compatible with the proposed model. Section 5 is a short summary.

1/25/00 11:11 AM

2 Conceptual Models of Information Seeking on the Web

Recent efforts to model information seeking on the Web have drawn upon metaphors and methods from fields as diverse as evolutionary biology and informetrics.

Just as animals evolve different methods of gathering and hunting food or prey in order to increase their intake of nutrition, humans also adopt different strategies of seeking information in order to increase their intake of knowledge. Foraging for information on the Web and foraging for food share common features: both resources tend to be unevenly distributed in the environment, uncertainty and risk characterize resource procurement, and all foragers are limited by time and opportunity costs as they choose to exploit one resource over another (Sandstrom 1994). Successful foragers are those who adopt strategies that maximize their harvest rates and their chances of survival. As a model in evolutionary biology, foraging theory requires some proxy currency as a measure of survival fitness. Since information does not deplete no matter how many have been 'feeding' on it, Sandstrom (1994) suggests that another characteristic of information, namely its novelty to the information seeker (and to his or her audience) be operationalized as a fitness currency. **Information foraging** refers to activities associated with assessing, seeking, and handling information sources, particularly in networked environments. Such search will be adaptive to the extent that it makes optimal use of knowledge about expected information value and expected costs of accessing and extracting the relevant information (Pirolli and Card 1995; Pirolli, Pitkow and Rao 1996). For example, a wolf hunts for prey, but a spider builds a web and waits for the prey to come to it. Humans seeking information also adopt different strategies, sometimes with close parallels to those of animal foragers. Pirolli and Card (1995) noted that the wolf-prey strategy resembles classic information retrieval, while the spider-web strategy is akin to information filtering. Pirolli, Pitkow and Rao (1996) suggest that the optimal selection of Web pages from a collection of related pages (a 'Web locality') to satisfy a user's information needs is a kind of optimal information diet problem. Optimality of the diet or pursuit sequence chosen by users will depend on their ability to rapidly categorize the Web page types, rank category members, assess their prevalences on the Web locality, assess the expected amount of return over cost of pursuit, and decide which categories to pursue and which to ignore.

Almind and Ingwersen (1997), Larson (1996), and Downie (1996), among others, have applied quantitative methods from **informetrics** to the Web. For example, Almind and Ingwersen (1997) regard the Web as a citation network where pages are the entities of information on the Web, with the hyperlinks from the pages acting as citations. They believe that the Web is well suited for informetric investigations of the links between Web information entities, because both the quoting entities and the quoted information are easily accessible. Furthermore, it is possible to carry out citation analysis by parsing HTML tags used to mark up Web pages. For example, the TITLE or first H1 tag may contain the document title; ADDRESS tag the author's name; EM or STRONG tags keywords; the URL the institutional affiliation, and so on. Almind and Ingwersen observe that "the possibilities available in citation indexes and full-text databases respectively can be combined on the WWW where it is possible to search a citation network that also contains the full texts." (pg. 406) The Web however presents its own special difficulties, as each author is free to mark up and thereby index his or her own information object (web page) according to individual preferences. Almind and Ingwersen have successfully applied methods similar to bibliometric analysis of citation databases to compare Denmark's proportion of the Web with those of other Nordic countries. Larson (1996) also successfully conducted an experiment applying cocitation analysis methods to produce "quite reasonable and comprehensible clusterings of WWW sites that had topical similarities" in the subject area of geographic information systems, earth sciences, and satellite remote sensing. Downie (1996) shows how informetric modelling techniques and principles may be used to analyse log files created by Web servers in order to reveal usage and interaction patterns at a Web site.

3 Towards a New Behavioral Model of Information Seeking on the Web

3 .1 Modes of Organizational Scanning

The models outlined in the last section are promising approaches, particularly in their ability to reveal global, historical patterns of use; suggest alternative metrics of information value; and provide implications for systems design. At the same time, models may also be needed that focus on the information behaviors of individuals as they traverse the Web, taking into account the context in which this information seeking is situated (addressing questions such as why was the information needed, and how was the information used). In this subsection, we review four modes of organizational scanning discussed in organization science. The next subsection (3.2) reviews a model based on six categories of information seeking activities. Subsection 3.3 combines elements from both models to propose a new behavioral framework for analyzing information seeking on the Web.

Research in organization science suggests that it might be helpful to distinguish between four modes of organizational scanning: undirected viewing, conditioned viewing, informal search, and formal search (Aguilar 1967, 1988; Weick and Daft 1983; Daft and Weick 1984).

In **undirected viewing**, the individual is exposed to information with no specific informational need in mind. The overall purpose is to scan broadly in order to detect signals of change early. Many and varied sources of information are used, and large amounts of information are screened. The granularity of information is coarse, but large chunks of information are quickly dropped from attention. The goal of broad scanning implies the use of a large number of different sources and different types of sources. These sources should supply up-to-date news and provide a variety of points of views. Information on the Web appears to match these requirements well. The Web is a laissez faire information marketplace offering a huge diversity of sources presenting information through a wide range of perspectives. Information often becomes available on the Web more quickly than through print channels. The immediacy, variety and eclecticism of the Web makes it a useful medium for detecting early, weak signals about trends and phenomena that could become significant over time. As a result of undirected viewing, general areas or topics may be identified as being potentially relevant to the organization's goals or tasks, and the individual becomes sensitive to these areas.

In **conditioned viewing**, the individual directs viewing to information about selected topics or to certain types of information. The overall purpose is to evaluate the significance of the information encountered in order to assess the general nature of the impact on the organization. The individual has isolated a number of areas of potential concern from undirected viewing, and is now sensitized to assess the significance of developments in those areas. The individual wishes to do this assessment in a cost-effective manner, without having to dedicate substantial time and effort in a formal search. The Web can provide a number of ways of obtaining information to make initial sense of emergent phenomena. For example, market research companies, financial institutions, industry associations, and government organizations make available on Web pages their reports, bulletins, and newsletters that analyze ongoing developments in their areas of watch. Some academics, authors, consultants, industry observers, and knowledgable experts use the Web to share their insights and predictions, and to stimulate further discussion. If the impact is assessed to be sufficiently significant, the scanning mode changes from scanning to searching.

During **informal search**, the individual actively looks for information to deepen the knowledge and understanding of a specific issue. It is informal in that it involves a relatively limited and unstructured effort. The overall purpose is to gather information to elaborate an issue so as to determine the need for action by the organization. The individual has determined the potential importance of specific developments, and embarks on a search that would build up knowledge about those developments, and deepen understanding of their implications and consequences. In conducting an informal search, the Web can address the requirement for information that is directed at specific issues, but that still does not cost a great deal of time or money to acquire. On the Web, search engines can be used to locate information on Web pages, newsgroups and mailing list discussions. Librarians and specialists have also compiled Web-based directories and lists of focused Web resources. If a need for a decision or response is perceived, the individual dedicates more time and resources to the search.

During **formal search**, the individual makes a deliberate or planned effort to obtain specific information

or information about a specific issue. Search is formal because it is structured according to some pre-established procedure or methodology. The granularity of information is fine, as search is relatively focused to find detailed information. The overall purpose is to systematically retrieve information relevant to an issue in order to provide a basis for developing a decision or course of action. Formal searches could be a part of for example, competitor intelligence gathering, patents searching, market demographics analysis, and issues management. Formal searches prefer information from sources that are perceived to be knowledgable, or from information systems and services that make efforts to ensure data quality and accuracy. The four modes of scanning are summarized and compared in Figure 1.

Scanning Modes (Aguilar 1967; Weick & Daft 1983)	Information Need	Information Use	Amount of Targeted Effort	Number of Sources	Tactics
Undirected Viewing	General areas of interest; specific need to be revealed	Serendipitous discovery "Browsing"	Minimal	Many	Scan broadly a diversity of sources, taking advantage of what's easily accessible "Visioning"
Conditioned Viewing	Able to recognize topics of interest	Increase knowledge about topics of interest "Learning"	Low	Few	Browse in pre-selected sources on pre-specified topics of interest "Discriminating"
Informal Search	Able to formulate queries	Increase knowledge on area within narrow boundaries "Selecting"	Medium	Few	Search is focused on area or topic, but a good-enough search (to solve a problem, move along) is satisfactory "Satisficing"
Formal Search	Able to specify targets	Formal use of information for decision-, policy-making "Retrieving"	High	Many	Systematic gathering of information about an entity, following some method or procedure "Optimizing"

Figure 1. Modes of Scanning

The individuals in an organization are simultaneously engaged in all four modes of scanning. They view the environment broadly in order to see the big picture as well as to identify areas that require closer attention. At the same time, they are searching for information on particular issues in order to assess their significance and to develop appropriate responses. Etzioni (1967, 1986) compares this "mixed scanning" to a satellite scanning the earth by using both a wide-angle and a zoom lens: "Mixed scanning ... is akin to scanning by satellites with two lenses: wide and zoom. Instead of taking a close look at all formations, a prohibitive task, or only at the spots of previous trouble, the wide lenses provide clues as to places to zoom in, looking for details." (Etzioni 1986, p. 8) Effective environmental scanning requires both general viewing that sweeps the horizon broadly and purposeful searching that probes issues in

6

sufficient detail to provide the kinds of information needed for decision making.

3.2 Ellis' Model of Information Seeking Behaviors

Ellis (1989), Ellis et al (1993), and Ellis and Haugan (1997) propose and elaborate a general model of information seeking behaviors based on studies of the information seeking patterns of social scientists, research physicists and chemists, and engineers and research scientists in an industrial firm. One version of the model describes six categories of information seeking activities as generic: starting, chaining, browsing, differentiating, monitoring, and extracting.

Starting comprises those activities that form the initial search for information -- identifying sources of interest that could serve as starting points of the search. Identified sources often include familiar sources that have been used before as well as less familiar sources that are expected to provide relevant information. The likelihood of a source being selected depends on the perceived accessibility of the source, as well as the perceived quality of the information from that source. Perceived accessibility, which is the amount of effort and time needed to make contact with and use a source, has been found to be a strong predictor of source use for many groups of information users (such as engineers and scientists (Allen 1977)). However, in situations when ambiguity is high and when information reliability is especially important, less accessible sources of perceived high quality may be consulted as well (see for example the environment scanning behavior of chief executives in Choo (1998)). While searching the initial sources, these sources are likely to point to, suggest, or recommend additional sources or references. Following up on these new leads from an initial source is the activity of **Chaining**. Chaining can be backward or forward. Backward chaining takes place when pointers or references from an initial source are followed, and is a well established routine of information seeking among scientists and researchers. In the reverse direction, forward chaining identifies and follows up on other sources that refer to an intial source or document. Although it can be an effective way of broadening a search, forward chaining is much less commonly used, probably because people are unaware of it or because the required bibliographical tools are unavailable.

Having located sources and documents, **Browsing** is the activity of semi-directed search in areas of potential search. The individual often simplifies browsing by looking through tables of contents, lists of titles, subject headings, names of organizations or persons, abstracts and summaries, and so on. Browsing takes place in many situations in which related information has been grouped together according to subject affinity, as when the user views displays at a conference or exhibition, or scans periodicals or books along the shelves of a bookshop or library. Chang and Rice (1993) define browsing as "the process of exposing oneself to a resource space by scanning its content (objects or representations) and/or structure, possibly resulting in awareness of unexpected or new content or paths in that resource space." (p. 258) They regard browsing as a "rich and fundamental human information behavior" that could lead to outcomes such as serendipitous findings, modification of information needs, learning, enjoyment, and so on. During **Differentiating**, the individual filters and selects from among the sources scanned by noticing differences between the nature and quality of the information offered. For example, social scientists were found to prioritize sources and types of sources according to three main criteria: by substantive topic; by approach or perspective; and by level, quality, or type of treatment (Ellis 1989). The differentiation process is likely to depend on the individual's prior or initial experiences with the sources, word-of-mouth recommendations from personal contacts, or reviews in published sources. Taylor (1986) points out that for information to be relevant and consequential, it should address not only the subject matter of the problem but also the particular circumstances that affect the resolution of that problem. He identifies six categories of criteria by which individuals select and differentiate between sources: ease of use, noise reduction, quality, adaptability, time savings, and cost savings.

Monitoring is the activity of keeping abreast of developments in an area by regularly following particular sources. The individual monitors by concentrating on a small number of what are perceived to be core sources. Core sources vary between professional groups, but usually include both key personal contacts and publications. For example, social scientists and physicists were found to track developments through core journals, online search updates, newspapers, conferences, magazines, books, catalogues, and so on (Ellis et al 1993). **Extracting** is the activity of systematically working through a

particular source or sources in order to identify material of interest. As a form of retrospective searching, extracting may be achieved by directly consulting the source, or by indirectly looking through bibliographies, indexes, or online databases. Retrospective searching tends to be labor intensive, and is more likely when there is a need for comprehensive or historical information on a topic.

Although the Ellis model is based on studies of academics and researchers, the categories of information seeking behaviors may be applicable to other groups of users as well. For example, Sutton's (1994) analysis of the information seeking behavior of attorneys noted that the three stages of legal research he identified (base-level modelling, context sensitive exploration, and disambiguating the space) could be mapped into Ellis's categories of starting, chaining, and differentiating. The identification of categories of information seeking behavior also suggests that information retrieval systems could increase their usefulness by including features that directly support these activities. Ellis thought that hypertext-based systems would have the capabilities to implement these functions (Ellis 1989). If we visualize the World Wide Web as a hyperlinked information system distributed over numerous networks, most of the information seeking behavior categories in Ellis' model are already being supported by capabilities available in common Web browser software. Thus, an individual could begin surfing the Web from one of a few favourite starting pages or sites (starting); follow hypertextual links to related information resources -- in both backward and forward linking directions (chaining); scan the Web pages of the sources selected (browsing); bookmark useful sources for future reference and visits (differentiating); subscribe to e-mail based services that alert the user of new information or developments (monitoring); and search a particular source or site for all information on that site on a particular topic (extracting). Plausible extensions of the acitivities to Web information seeking (labelled Web Moves), are compared with the original formulations (Literature Search Moves) in Figure 2 below.

Information Seeking Behaviors (Ellis and Ellis et al, 1989, 1993, 1997)	Starting	Chaining	Browsing	Differentiating	Monitoring	Extractin
Literature Search Moves	Identifying sources of interest	Following up references found in given material	Scanning tables of contents or headings	Assessing or restricting information according to their usefulness	Receiving regular reports or summaries from selected sources	Systematica working through a source to identify material of interest
Anticipated Web Moves	Identifying websites/ pages containing or pointing to information of interest	Following links on starting pages to other content-related sites	Scanning top-level pages: lists, headings, site maps	Selecting useful pages and sites by bookmarking, printing, copying and pasting, etc Choosing/starting at differentiated, pre-selected site of known content	Receiving site updates using e.g. push, agents, or profiles Revisiting 'favorite' sites for new information	Systematica searches a local site to extract information interest at th site

Figure 2. Information Seeking Behaviors and Web Moves

3.3 Towards a New Behavioral Model of Information Seeking on the Web

Aguilar's modes of scanning and Ellis's seeking behaviors may be combined and extended in a new behavioral model of information seeking on the Web. The figure below identifies four main modes of information seeking on the Web: undirected viewing, conditioned viewing, informal search, and formal search. For each mode, the figure indicates which information seeking activities or moves are likely to dominate, as suggested by theory.

	Starting	Chaining	Browsing	Differentiating	Monitoring	Extracting
Undirected Viewing	✓	✓				
Conditioned Viewing			✓	✓	✓	
Informal Search				✓	✓	✓
Formal Search					✓	✓

Figure 3. Behavioral Model of Information Seeking on the Web

3.3.1 Undirected Viewing

In the undirected viewing mode, while there are broad areas of interest, there is no particular information need that may be articulated explicitly or formally. Instead, the purpose of viewing is precisely to notice significant developments or issues that then generate new information needs. As noted earlier, typical tactics here would involve viewing a diversity of sources, taking advantage of what's easily accessible, and including sources which may not seem at first to be directly related to the work of the organization.

In terms of information seeking moves on the Web, we may anticipate starting and chaining to dominate. Starting occurs when viewers begin their web use on pre-selected default home pages, or when they visit a favorite page or site to begin their viewing (such as news, newspaper, or magazine sites). Chaining occurs when viewers notice items of interest (often by chance), and then follow hypertext links to more information on those items. Forward chaining of the sort just described is the most typical during undirected viewing. Backward chaining is also possible, since search engines can be used to locate other Web pages that point to the site that the user is currently at.

3.3.2 Conditioned Viewing

In the conditioned viewing mode, there are specific topic areas that define the scope and substance of the viewer's information needs. The viewer is sensitive to information about these topics, and is able to assess, in a general way, the significance of the information encountered. To increase knowledge on these topics, typical tactics would involve browsing in sources that the viewer knows to contain potentially useful information.

In terms of information seeking moves on the Web, we may anticipate browsing, differentiating, and monitoring to be common. Differentiating occurs as viewers select Web sites or pages that they expect to provide relevant information. Sites may be differentiated based on prior personal visits, or recommendations by others (such as word-of-mouth or published reviews). Differentiated sites are often bookmarked. When visiting differentiated sites, viewers browse the content by looking through tables of contents, site maps, or list of items and categories. Viewers may also monitor highly differentiated sites

by returning regularly to browse, or by keeping abreast of new content (through, for example subscribing to newsletters that report new material on the site).

3.3.3 Informal Search

During informal search, the individual has amassed enough knowledge and awareness about a topic to formulate a query to learn more about a specific issue or development. An informal search query is possible because the individual is able to establish some parameters and boundaries to constrain the search. At the same time, the search is limited as the individual does not wish to expend substantial amounts of time and effort. The purpose is to learn more about the issue in order to determine the need for action or response.

In terms of moves on the Web, we may anticipate differentiating, extracting, and monitoring to be typical. Again, informal search is likely to be attempted at a small number of Web sites that have been differentiated by the individual, based on the individual's knowledge about these sites' information relevance, quality, affiliation, dependability, and so on. Extracting is relatively "informal" in the sense that searching would be localized to looking for information within the selected site(s). Extracting is also likely to make use of the basic, 'simple' search features or commands of the local search engine, in order to get at the most important or most recent information, without attempting to be comprehensive. Monitoring becomes more proactive if the individual sets up push channels or software agents that automatically find and deliver information based on selection of keywords or topics.

3.3.4 Formal Search

During formal search, the individual is prepared to invest substantial time and effort in order to gather information that will enable action to be taken. The search may be formal because it follows some pre-established routine or method. The search is also formal because it is now possible, (with the knowledge from informal search and conditioned viewing,) to elaborate the query in detail -- specifying the target of inquiry or retrieval according to desired attributes (authors, institutions, dates, document types, and so on). Information gained from formal search is typically used 'formally' as well, for policy making, strategic planning, and other forms of decision making.

In terms of moves on the Web, we may anticipate primarily extracting operations, with some complementary monitoring activity. Formal search makes use of search engines that cover the Web relatively comprehensively, and that provide a powerful set of search features that can focus retrieval. Because the individual wishes not to miss any important information, there is a willingness to spend more time in the search, to learn and use complex search features, and to evaluate the sources that are found in terms of quality or accuracy. Formal search may be two-staged: multi-site searching that identify significant sources is then followed by within-site searching. Within-site searching may involve fairly intensive foraging. Extracting may be supported by monitoring activity, again through services such as Web site alerts, push channels, and software agents, in order to keep up with late-breaking information.

4 The Pilot Study

4.1 Research Design

This paper presents findings from a pilot field study to investigate the information seeking behaviors of Web users. The behavioral model presented in this paper emerged as much from an analysis of the field data collected as from a synthesis of theoretical concepts in information science and organization science. Phase 2 of the study is in progress at the time of writing (Spring 1998), and at the end of both phases, a total of 30 individuals would have participated in the study. Participants were selected according to the general criterion that they employ the Web routinely to find and use information for their work-related needs. The study sample included a number of managers, IT specialists, and information specialists.

Eleven persons took part in the pilot study. Three are managers working in very large corporations (an international bank, and a utility company); three are IT architects; two are technology consultants; two are research and technical support specialists, and one is president of his own software firm. Nine of the participants are very knowledgable about IT. Although the number of participants was small, their Web use behaviors were monitored continuously over two-week periods. The unit of analysis was thus the individual information seeking episode, and the relatively fine-grained data collection and analysis provided a useful first iteration of testing the conceptual model developed in this paper. .

4.2 Data Collection

Four methods of data collection were employed: questionnaire survey; tracker application that recorded Web browser actions; proxy server that logged Web resource and service requests; and personal interviews with participants.

The **questionnaire survey** was administered at the participants' work places, during the first visit. The survey contained 12 questions that identifed the information sources the participants used, their frequency of using these sources, and their perception of the perceived accessibility and quality of each of the sources. A wide range of sources was covered, including personal and impersonal sources (print and electronic), as well as internal and external sources. There were also questions on the amount of time and frequency of using the Web for information seeking. Furthermore, through informal conversations during the visit, research team members were able to develop a general impression of the style and scope of each participant's Web use.

The **Tracker application** was specially designed and developed for this study. The Tracker was installed on each participant's computer, and it ran transparently whenever the participant's Web browser was being used. The Tracker application was left to run on participants' computers for two-week periods. Because the Tracker was essentially 'invisible,' it was not expected to influence participants' normal Web-use behaviors. After two weeks, Tracker was uninstalled, and the Tracker log file collected for analysis. The Tracker recorded how each participant was using the browser to navigate the Web and manipulate information from the Web. Specifically, it recorded all URL calls and requests, as well as most browser menu selections, and wrote these events into a local log file on each participant's hard disk. Browser menu selections captured included "Open URL or File," "Reload," "Back," "Forward," "Add to Bookmarks," "Go to Bookmark," "Print," and "Stop." Because all URL calls and menu selections were date-time stamped as they were written into the Tracker log, the research team was able to subsequently reconstruct move-by-move how participants looked for information on the Web during particular episodes.

For a few sites where the Tracker application was not usable, a Web **proxy server** was set up to collect data on what sites and pages were accessed by participants. The settings of each participant's Web browser were changed to redirect all HTTP requests to a proxy server monitored by the research team in the University of Toronto. The proxy server's transfer log recorded the IP address of the participant requesting a file, the date and time of the transfer, the HTTP method and protocol used for the transfer, the status of the transfer, and how many bytes were transferred. The proxy log recorded the full URL addresses of all files requested, as well as any variables that were sent along with the URL. The latter provided important data on arguments and attributes that were sent along to search engines and other back-end applications at remote host sites. As with the Tracker, the use of the proxy server was transparent to participants, and the use of a fast proxy ensured that there was imperceptible performance impact on transferring files. At the end of the two-week period, the proxy server setting was deleted in participants' browser software.

The event log and transfer log from the Tracker application and the proxy server were pre-analyzed to prepare for **personal interviews** that were conducted with each participant. The interview format was based on the principles of the Critical Incident Technique (Flanagan 1954), in which the 'incident' to be studied should be recent, sufficiently complete, and its effects or consequences sufficiently clear. In the interviews, participants described two 'critical incidents' of Web information seeking and use in reply to the following question:

1/25/00 11:11 AM

"Please try to recall a recent instance in which you found important information on the Web, information that led to some significant action or decision. Would you please describe that incident for me in enough detail so that I can visualize the situation?"

Where appropriate, participants were prompted with the names of Web sites that were indicated in their Tracker or proxy log files. Besides 'critical incidents,' participants were also invited to comment more broadly on their use of the Web, including their general Web-use strategies and preferences, as well as what they perceived to be both the positive and negative aspects of Web use.

4.3 Data Analysis

The Tracker and/or proxy server log files were tabulated into large spreadsheets with entries arranged in chronological sequence. Each entry contained a date-time value, followed by a URL or a browser menu action name. Entries were grouped into major clusters indicating extended or frequent visits to particular Web sites. The log tables were then re-examined together with data from the personal interviews in order to identify "significant" episodes of information seeking for further analysis. The selection of episodes was guided by

- a highlighting of the episode by the participant during the personal interview;
- evidence of the episode having consumed a relatively substantial amount of time and effort;
- evidence that the episode was a recurrent activity.

Each significant episode of information seeking was then classified according to the **mode** of scanning or information seeking, and the **moves** that were employed in that mode. Where available, interview data helped determine the **mode** of scanning or information seeking. Using the behavioral model presented earlier and summarized in Figure 1, participants' verbal descriptions of the context, information needs, information use, and amount of effort were analyzed to infer whether the mode was undirected viewing, conditioned viewing, informal search, or formal search. Data from Tracker and proxy server log files helped determine the **moves** exercised by participants as they use their Web browsers to view and find information. Data about the sequence of site visits, repetitions of these sequences, movements backwards and forwards between pages, the use of bookmarking, the selection of sites from stored bookmarks, the use of search engines, printing, and other actions and events captured by the Tracker and proxy logs were examined to trace the selection and development of information seeking moves over the duration of each episode. Using the criteria presented earlier (based on Ellis' model) and summarized in Figure 2, participants' moves were analyzed to infer whether moves may be classified as starting, chaining, browsing, differentiating, monitoring, or extracting.

4.4 Preliminary Results and Discussion

Thirty episodes of 'significant' information seeking were identified and classified according to the modes and moves of information seeking defined in Section 5. The majority of the episodes were classified as informal search (11) and conditioned viewing modes (10). A smaller number of episodes were undirected viewing (5) and formal search (4). Figure 4 below shows the distribution of the episodes over the four modes of viewing and searching.

	Starting	Chaining	Browsing	Differentiating	Monitoring	Extracting
Undirected Viewing	5 Episodes					
Conditioned Viewing			10 Episodes			
Informal Search				11 Episodes		
Formal Search						4 Episodes

Figure 4. Episodes of Information Seeking on the Web

The episodes in each mode were examined in terms of their Web moves. In the undirected viewing episodes, data collected by the Tracker application and/or proxy server suggested that the main moves were starting (beginning at a favorite jumpsite) and chaining (following links on that site). In the conditioned viewing episodes, the main moves appeared to be differentiating (selecting known, recommended or bookmarked sites; printing selected pages), browsing (scanning top-level pages, table of contents, site maps), and monitoring (revisiting favorite sites regularly in order to check for new or updated content). In the informal search episodes, the main moves observed were localized extracting (using search engines dedicated to retrieving information from the local site), differentiating (pre-selecting sites to search in, printing pages), and monitoring (regular return visits). In the formal search episodes, the main move was more intensive and careful extracting, involving the use of search engine(s) which indexed numerous sites and/or historical data, and the retrieval of multiple items addressing the same information need. Table 1 shows two example episodes in each mode, as well as the Web moves enacted.

The data appear to be compatible with the behavioral model of Web information seeking developed in this paper (compare the empirical observations in Table 1 with predicted Web moves in Figure 3). Thus, the model's four modes of viewing and searching seems to be a feasible and useful method of distinguishing between different modes of information seeking on the Web. These modes were in turn set apart by their context (information needs), purpose (information use), and scope (amount of effort and number of sites).

Moreover, the model's predictions about the likely moves for browsing and finding information on the Web within each of the viewing/searching modes seem to have been largely borne out by the empirical data collected. Thus, undirected viewing was mainly characterized by starting and chaining; conditioned viewing by differentiating, browsing, and monitoring; informal search by differentiating, extracting, and monitoring; and formal search by relatively in-depth, careful extracting.

While there was broad overlap between predicted and observed Web moves, there were also a few interesting divergences. Most of the information seeking episodes were in the modes of conditioned viewing and informal search. There were only a few episodes of information seeking in the formal search mode. When they did occur, formal search operations were only incrementally more sophisticated than those in informal searches.

Most instances of **monitoring** moves were in the form of regular return visits to sites which the participants knew would contain useful information that would be updated. Although most participants were relatively savvy Web users, only a few of them took advantage of advanced methods to keep up with new content. One used an e-mail alert service, three others subscribed to a push service (but all three subsequently uninstalled it).

Most instances of **extracting** also employed straightforward retrieval methods. This was the case even when participants appeared to be working in the formal search mode. (As noted earlier, formal searches

were only marginally more intricate than informal searches.) For the most part, search formulations were relatively simple, with advanced features such as Boolean operators, and word truncation or proximity operators rarely utilized.

Mode	Label	Episode		ST	CH	BR	DI	MO	EX
Undirected Viewing	F3	Starting from news.com page, followed links to items of interest; including clicking on banner ads.	✓	✓					
	H1	Found by accident, site with demo of one-handed typing application, followed links to other related sites.	✓	✓					
Conditioned Viewing	D1	Knew of IJC site as good source for info on copper and zinc emissions, viewed site, and showed site to colleagues on intergovernmental committee.			✓	✓			
	F4	Visited Sun's Java Home Page, viewed pages on Electronic Commerce Framework, and E-com Tools.			✓	✓	✓		
Informal Search	J4	From personalized Forrester page, and using Forrester's search engine, retrieved company reports, and printed one.				✓	✓	✓	
	R2	At Microsoft site, used local search tool to find specs on library functions, this solved a technical programming problem.				✓	✓	✓	
Formal Search	A1	Used search engine to look for formal definition of "model view controller." Found 5 good definitions, discussed them with colleagues, used in technical documentation.						✓	
	D3	Used DejaNews search engine to retrieve 2 author profiles and 6 documents.						✓	
ST=Starting; **CH**=Chaining; **BR**=Browsing; **DI**=Differentiating; **MO**=Monitoring; **EX**=Extracting									

Table 1. Examples of Information Seeking Episodes

Aggregated data from the questionnaire surveys and personal interviews showed a number of interesting patterns:

- Of the twelve information sources that were compared, the Web was the third most frequently used source, after colleagues and mass media.
- On average, participants spent about 20% of their work hours on the Web.
- The majority of participants were using the Web to look for technical information.
- Quality of information from the Web was perceived to be very high, surpassed only by one other source: "colleagues in the same department." However, some critical comments were raised during the personal interviews.
- Human sources were still valued most highly: colleagues in same department were rated as providing information of the highest quality.
- The Web as a source was perceived to be as accessible as other "internal" information sources such as managers and supervisors, internal memos, and other colleagues. However, the Web was seen as being less accessible than mass media sources such as radio and television.
- Few participants deliberately set out to search for new sites; instead sites visited were recommended from other sources, or they simply stumbled across good sites.

5 Summary

The research presented here developed a new behavioral model of information seeking on the Web by combining theoretical elements from information science and organization science. Specifically, the model was constructed by distinguishing between four modes of organizational scanning (undirected viewing, conditioned viewing, informal search, formal search), and six generic moves of information seeking (starting, chaining, browsing, differentiating, monitoring, extracting). The model was then tested in a pilot study which collected and analyzed data on how a sample of participants in their natural work settings sought and used information from the Web. Findings from the pilot study appeared to support the behavioral model, both in terms of the modes of scanning, and the moves of Web information seeking associated with each mode. Overall, the study suggests that a behavioral framework that relates motivations (the strategies and reasons for viewing and searching) and moves (the tactics used to find and use information) may be helpful in analysing Web-based information seeking. The study also suggests that multiple, complementary methods of collecting qualitative and quantitative data may be used within a single study to compose a richer portrayal of how individuals seek and use Web-based information in their natural work settings. This paper presents results from Phase 1 of an ongoing larger study -- it is hoped that results from Phase 2 would be presented in a future ASIS Meeting.

References

Aguilar, Francis J. 1967. Scanning the Business Environment. New York, NY: Macmillan Co.

Aguilar, Francis J. 1988. General Managers in Action. New York, NY: Oxford University Press.

Allen, Thomas J. 1977. Managing the Flow of Technology: Technology Transfer and the Dissemination of Technological Information within the R & D Organization. Cambridge, MA: MIT Press.

Almind, Tomas C. and Peter Ingwersen. 1997. Infometric Analysis on the World Wide Web:

Methodological Approaches to "Webometrics". Journal of Documentation 53, no. 4: 404-426.

Chang, Shan-Ju and Ronald E. Rice. 1993. Browsing: A Multidimensional Framework. In Annual Review of Information Science and Technology , ed. Martha E. Williams. Medford, NJ: Learned Information.

Choo, Chun Wei. 1998. Information Management for the Intelligent Organization: The Art of Scanning the Environment. Second ed. Medford, NJ: Information Today, Inc.

Daft, Richard L. and Karl E. Weick. 1984. Toward a Model of Organizations as Interpretation Systems. Academy of Management Review 9, no. 2: 284-295.

Downie, J. Stephen. 1996. Informetrics and the World Wide Web: A Case Study and Discussion. In Proceedings of Canadian Association for Information Science held in University of Toronto, Toronto, Ontario, edited by Charles Meadow et al, p. 130-141. Canadian Association for Information Science.

Ellis, David and Merete Haugan. 1997. Modelling the Information Seeking Patterns of Engineers and Research Scientists in an Industrial Environment. Journal of Documentation 53, no. 4: 384-403.

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

Ellis, David. 1989. A Behavioural Model for Information Retrieval System Design. Journal of Information Science 15, no. 4/5: 237-247.

Etzioni, Amitai. 1967. Mixed-Scanning: A "Third" Approach to Decision-Making. Public Administration Review 27, no. 5:385-392.

Flanagan, John C. 1954. The Critical Incident Technique. Psychological Bulletin 51, no. 4: 327-358.

Larson, Ray R. 1996. Bibliometrics of the World Wide Web: An Exploratory Analysis of the Intellectual Structure of Cyberspace. In Proceedings of 59th ASIS Annual Meeting held in Baltimore, Maryland, edited by Steve Hardin, Vol. 33: 71-78. Information Today Inc.

Pirolli, Peter and Stuart Card. 1995. Information Foraging in Information Access Environments. In Proceedings of Conference on Human Factors in Computer Systems, CHI-95 held in Denver, Colorado, USA, p. 51-58. ACM Press.

Pirolli, Peter, James Pitkow, and Ramana Rao. 1996. Silk from a Sow's Ear: Extracting Usable Structures from the Web. In Proceedings of Conference on Human Factors in Computer Systems, CHI-96 held in Vancouver, BC, Canada. ACM Press.

Sandstrom, Pamela Effrein. 1994. An Optimal Foraging Approach to Information Seeking and Use. Library Quarterly 64, no. 4: 414-449.

Weick, Karl E. and Richard L. Daft. 1983. The Effectiveness of Interpretation Systems. In Organizational Effectiveness: A Comparison of Multiple Models, ed. Kim S. Cameron and David A. Whetten, 71-93. New York, NY: Academic Press.

1/25/00 11:11 AM

REPRODUCTION RELEASE

(Specific Document)

I. DOCUMENT IDENTIFICATION:

Title:

A Behavioral Model of Information Seeking on the Web-Preliminary Results of a
Study of How Managers and IT Specialists Use the Web

Author(s): Chun Wei Choo, Brian Detlor, Don Turnbull

Corporate Source:	Publication Date: 1998

II. REPRODUCTION RELEASE:

In order to disseminate as widely as possible timely and significant materials of interest to the educational community, documents announced in the monthly abstract journal of the ERIC system, *Resources in Education* (RIE), are usually made available to users in microfiche, reproduced paper copy, and electronic media, and sold through the ERIC Document Reproduction Service (EDRS). Credit is given to the source of each document, and, if reproduction release is granted, one of the following notices is affixed to each document.

If permission is granted to reproduce and disseminate the identified documents, please CHECK ONE of the following three options and sign at the bottom of the page.

The sample sticker shown below will be affixed to all Level 1 documents	The sample sticker shown below will be affixed to all Level 2A documents	The sample sticker shown below will be affixed to all Level 2B documents
PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY ___ Sample ___ TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) 1	PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE, AND IN ELECTRONIC MEDIA FOR ERIC COLLECTION SUBSCRIBERS ONLY, HAS BEEN GRANTED BY ___ Sample ___ TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) 2A	PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE ONLY HAS BEEN GRANTED BY ___ Sample ___ TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) 2B
Level 1 ↑ [✓]	Level 2A ↑ [ ]	Level 2B ↑ [ ]
Check here for Level 1 release, permitting reproduction and dissemination in microfiche or other ERIC archival media (e.g., electronic) and paper copy.	Check here for Level 2A release, permitting reproduction and dissemination in microfiche and in electronic media for ERIC archival collection subscribers only	Check here for Level 2B release, permitting reproduction and dissemination in microfiche only

Documents will be processed as indicated provided reproduction quality permits.
If permission to reproduce is granted, but no box is checked, documents will be processed at Level 1.

I hereby grant to the Educational Resources Information Center (ERIC) nonexclusive permission to reproduce and disseminate these documents as indicated above. Reproduction from the ERIC microfiche or electronic media by persons other than ERIC employees and its system contractors requires permission from the copyright holder. Exception is made for non-profit reproduction by libraries and other service agencies to satisfy information needs of educators in response to discrete inquiries.

Sign here→ please	Signature: *Brian Detlor*	Printed Name/Position/Title: BRIAN DETLOR, DOCTORAL CANDIDATE	
	Organization/Address: U of T	Telephone:	FAX:
		E-Mail Address: detlor@fis.utoronto.ca	Date: 1/26/00

III. DOCUMENT AVAILABILITY INFORMATION (FROM NON-ERIC SOURCE):

If permission to reproduce is not granted to ERIC, or, if you wish ERIC to cite the availability of the document from another source, please provide the following information regarding the availability of the document. (ERIC will not announce a document unless it is publicly available, and a dependable source can be specified. Contributors should also be aware that ERIC selection criteria are significantly more stringent for documents that cannot be made available through EDRS.)

Publisher/Distributor:
Address:
Price:

IV. REFERRAL OF ERIC TO COPYRIGHT/REPRODUCTION RIGHTS HOLDER:

If the right to grant this reproduction release is held by someone other than the addressee, please provide the appropriate name and address:

Name:
Address:

V. WHERE TO SEND THIS FORM:

Send this form to the following ERIC Clearinghouse:	ERIC Clearinghouse on Information & Technology Syracuse University 621 Skytop Road, Suite 160 Syracuse, NY 13244-5290 E-Mail eric@ericir.syr.edu 315-443-3640 1-800-464-9107 Fax: 315-443-5448

However, if solicited by the ERIC Facility, or if making an unsolicited contribution to ERIC, return this form (and the document being contributed) to:

EFF-088 (Rev. 9/97)

ERIC