
Information seeking on the Web by women in IT professions

*Chun Wei Choo and
Christine Marton*

The authors

Chun Wei Choo is a Professor in the Faculty of Information Studies, University of Toronto, Toronto, Ontario, Canada.

Christine Marton is at the School of Library, Archival and Information Studies, The University of British Columbia, Vancouver, British Columbia, Canada.

Keywords

Generation and dissemination of information, Worldwide web, Women, Internet, Information network, Information gathering

Abstract

The paper develops a behavioral model of Web information seeking that identifies four complementary modes of information seeking: undirected viewing, conditioned viewing, informal search, and formal search. In each mode of viewing or searching, users would adopt distinctive patterns of browser moves: starting, chaining, browsing, differentiating, monitoring, and extracting. The model is applied empirically to analyze the Web information seeking behavior of 24 women in IT professions over a two-week period. Our results show that participants engaged in all four modes of information seeking on the Web, and that each mode may be characterized by certain browser actions. Overall, the study suggests that a behavioral approach that links information seeking modes (goals and reasons for browsing and searching) to moves (actions used to find and view information) may be helpful in understanding Web-based information seeking.

Electronic access

The Emerald Research Register for this journal is available at <http://www.emeraldinsight.com/researchregister>

The current issue and full text archive of this journal is available at

<http://www.emeraldinsight.com/1066-2243.htm>

Internet Research: Electronic Networking Applications and Policy
Volume 13 · Number 4 · 2003 · pp. 267-280
© MCB UP Limited · ISSN 1066-2243
DOI 10.1108/10662240310488951

Introduction

This paper presents findings of a research study on how women in IT professions seek information on the Web as part of their daily work. We begin with the premise that in order to analyze information seeking, we need to examine the context of the overall information task. That is, we need to understand the information needs, goals, and intended information use that motivate the information seeking in the first place. Furthermore, since we are interested in information seeking on the Web, we take advantage of the Web's ability to generate data about itself and how it is being used. Specifically, we analyze log files that capture data about how users employ their Web browsers to look for information. Our research is driven by the belief that a more complete understanding of information seeking behavior may be gained by linking the general contexts of information seeking tasks with the detailed sequences of Web-based actions and events that constitute the information seeking itself.

Research on information seeking on the Web

Catledge and Pitkow (1995) modified the source code for a version of XMosaic to generate a client-side log file that showed user navigation strategies and interface selections. They determined session boundaries by measuring the time between each event for all events, and adopted the heuristic that a lapse of 25.5 minutes or greater indicated the end of a "session." This heuristic is currently the most commonly used for delimiting sessions. The study found that Web pages users bookmarked did not match the most popular sites visited as a whole from the group. Only 2 percent of Web pages were either saved locally or printed. Catledge and Pitkow (1995) also hypothesized that users in their study categorized as "browsers" spent less time on a Web page than "searchers."

This research is supported by a grant from the Social Sciences and Humanities Research Council of Canada.



Tauscher and Greenberg (1997a, b) focused on the history mechanisms that Web browsers use to manage recently requested Web pages, and also modified the XMosaic browser to collect Web browsing data for over six weeks, from 23 participants. They found that 58 percent of the pages visited during a Web browsing session were re-visits. Users not only accessed a few pages frequently (60 percent once, and 19 percent twice), they also browsed in very small clusters of pages. Thus, Web browsing is a “recurring system . . . where users predominantly repeat activities they had invoked before, while still selecting new actions from the many that are possible” (Tauscher and Greenberg, 1997b, p. 400). People explained that they revisited Web pages because “the information contained by them changes; they wish to explore the page further; the page has a special purpose (e.g. search engine, home page); they are authoring a page; or the page is on a path to another revisited page” (Tauscher and Greenberg, 1997b, p. 400).

In a well known study, Huberman *et al.* (1998) analyzed AOL (America Online) usage data collected over a five-day period, and found a mathematical “law of surfing, . . . that determines the probability distribution of the depth - that is, the number of pages a user visits within a Web site” (Huberman *et al.*, 1998, p. 95). They started with a probabilistic model of the number of links a user might follow on a Web site. Next they calculated a value for the current page and related this to the value of the next page to be accessed. When the cost of moving to the next Web page is more than its expected value, the user stops Web surfing. Huberman *et al.* (1998) also examined Web server logs of the Xerox external Web site in order to obtain a bounded set of Web page requests. Cookies were used to track the paths of individual users as they surfed the site. By applying the surfing model with a spreading activation algorithm, they were able to predict the number of requests for each Web page in a Web site.

Hölscher and Strube (2000) investigated the effects of domain knowledge and Web knowledge on searching behavior and outcomes. Through two experiments, they found that participants who could rely on both types of expertise were overall most successful

in their searches. Participants strong in domain knowledge but lacking in Web searching knowledge relied heavily on terminology and avoided the use of search operators and modifiers. Participants with lower levels of knowledge were less flexible in their search strategies and tended to return to earlier stages of their searches rather than trying new approaches.

Cockburn and McKenzie (2001) analyzed four months of client-side log data that recorded user actions with Netscape Navigator, including page title, URL and time of each page visit, how often they visited each page, how long they spent at each page, the growth and content of bookmark collections, and other aspects of user interaction. The results showed that:

- revisiting Web pages is a much more prevalent activity than previously reported (approximately 81 percent of pages have been previously visited by the user);
- most pages are visited for a surprisingly short period of time;
- users maintain large (and possibly overwhelming) bookmark collections; and
- there is a marked lack of commonality in the pages visited by different users.

Cothey (2002) analyzed the real world Web information searching behavior of 206 college students over a ten-month period. The study used a longitudinal transaction log analysis of the URLs accessed during 5,431 user days of Web information searching to detect changes in information searching behavior associated with increased experience of using the Web.

Contrary to expectations, the study found that as users became more experienced, they adopted a more passive or browsing approach to Web information searching, and grew more eclectic in their selection of Web hosts.

Kim and Allen (2002) investigated the effects of task, cognitive style, and search engine on searching activities and outcomes. Two experiments generated nearly 160 completed searches by 80 individuals. The study found that of the three variables, only task showed a significant main effect. (Participants were assigned to a “known-item” search task and a “subject” search task.) Moreover, task interacted significantly with both search engine and cognitive abilities to influence how users

searched the Web. Thus, although the flexibility of the Web search engines allows different users to complete different search tasks successfully, search activities and outcomes were strongly influenced by task effects, and the techniques used and the efficiency of searches depend on how well the individual searcher fits with the specific task.

Towards a new behavioral model of information seeking on the Web

Modes of browsing and searching

Marchionini (1995) reviewed the research on browsing and observed that “there seems to be agreement on three general types of browsing that may be differentiated by the object of search (the information needed) and by the systematicity of tactics used” (Marchionini, 1995, p. 106). Directed browsing occurs when browsing is systematic, focused, and directed by a specific object or target. Semi-directed browsing occurs when browsing is predictive or generally purposeful: the target is less definite and browsing is less systematic. Finally, undirected browsing occurs when there is no real goal and very little focus. In a similar vein, Wilson (1997) identifies four categories of information seeking and acquisition: passive attention, where there may be no information seeking intended, but information acquisition may take place nevertheless; passive search, when one type of search results in the acquisition of information that happens to be relevant to the individual; active search, where an individual actively seeks out information; and ongoing search, where active searching has already established the basic framework of ideas, beliefs, but occasional continuing search is carried out to update or expand one’s framework.

In organization science, a comparable categorization of modes of organizational scanning or “browsing” has been developed, based on both empirical and theoretical research. The initial field work of Aguilar (1967) and the subsequent theoretical expansion by Weick and Daft (1983) suggest that organizations scan in four distinct modes: undirected viewing, conditioned viewing, informal search, formal search (Table I). 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 browsed. In conditioned viewing, the individual directs viewing to information about selected topics or to certain types of information. 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. 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. During formal search, the individual makes a deliberate or planned effort to obtain specific information or types of information about a particular issue. Search is formal because it is structured according to some procedure or methodology. The overall purpose is to systematically retrieve information in order to provide a basis for developing a decision or course of action.

Ellis’ model of information seeking behaviors

Ellis (1989), Ellis *et al.* (1993), and Ellis and Haugan (1997) developed a general model of information seeking behaviors based on 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. While viewing the initial sources, they 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. Having located sources and documents, browsing is semi-directed viewing in areas of potential search. The individual often simplifies browsing by looking through tables of contents, lists of titles and names, abstracts and summaries.

Table I Modes of scanning

Scanning modes	Information need	Information seeking	Information use
Undirected viewing	General areas of interest; specific need to be revealed	Scan broadly a diversity of sources, taking advantage of what is easily accessible	Serendipitous discovery "browsing"
Conditioned viewing	Able to recognize topics of interest	Browse in pre-selected sources on pre-specified topics of interest	Increase knowledge about topics of interest "learning"
Informal search	Able to formulate simple queries	Search is focused on area or topic, but a good-enough search is satisfactory	Increase knowledge on area within narrow boundaries "satisficing"
Formal search	Able to specify targets in detail	Systematic gathering of information about an entity, following some method or procedure	Formal use of information for decision-, policy-making "retrieving"

During differentiating, the individual selects and prioritizes from among the sources scanned by noticing differences between the nature and quality of the information offered. Monitoring is keeping abreast of developments in an area by regularly following a small number of what are perceived to be core sources. Extracting is systematically working through a particular source or sources in order to identify material of interest. Retrospective searching is labor intensive, and is more likely when there is a need for comprehensive or historical information on a topic.

Ellis (1989) thought that hypertext-based systems would have the capabilities to implement functions indicated by his behavioral model. Indeed, most of the information seeking behavior categories in Ellis' (1989) model are already being supported by common Web browser software. Thus, an individual could begin surfing the Web from one of a few favorite pages or sites (starting); follow 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 activities to Web information seeking (labeled

Web moves), are compared with the original formulations (literature search moves) in Table II.

Towards a behavioral model of information seeking on the Web

Aguilar's (1967) modes of scanning and Ellis's (1989) seeking behaviors may be combined in a behavioral model of information seeking on the Web. The model hypothesizes that there are four main modes of information seeking on the Web: undirected viewing, conditioned viewing, informal search, and formal search (Table III). For each mode, certain browser actions are likely to occur more frequently, based on our extension of the work of Ellis (1989).

In the undirected viewing mode on the Web, we may expect to see many instances of starting and chaining. Starting occurs when people begin their Web use at general information or omnibus sites to start their viewing (such as news, newspaper, or magazine sites). Chaining occurs when people notice items of interest, and follow links to more information on those items. In the conditioned viewing mode on the Web, we expect 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. Differentiated sites may be bookmarked. At these sites, users browse by looking through

Table II Information seeking behaviors and Web moves

	Starting	Chaining	Browsing	Differentiating	Monitoring	Extracting
Literature search moves (Ellis et al., 1993)	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	Systematically working a source to identify material of interest
Anticipated Web moves	Identifying Web sites 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 by bookmarking, printing, copying, etc. Choosing differentiated, pre-selected site	Receiving site updates via alerts or subscriptions. Revisiting "favorite" sites	Systematically searches a local site to extract information of interest at that site

Table III Behavioral modes and moves of information seeking on the Web

	Starting	Chaining	Browsing	Differentiating	Monitoring	Extracting
Undirected viewing	Identifying selecting starting pages, sites	Following links on initial pages				
Conditioned viewing			Browsing entry pages, headings, site maps	Bookmarking, printing, copying Going directly to known site	Revisiting "favorite" or bookmarked sites; alerts, subscriptions	
Informal search				Bookmarking, printing, copying Going directly to known site	Revisiting "favorite" or bookmarked sites; alerts, subscriptions	Using (local) search engines to extract information
Formal search					Revisiting "favorite" or bookmarked sites; alerts, subscriptions	Using search engines to extract information

tables of contents, site maps, or list of items and categories. Users may also monitor differentiated sites by returning regularly, or by signing up for alerts. During informal search on the Web, we expect differentiating and extracting to be typical. Extracting is relatively "informal" in that it makes use of keywords and basic search features, sometimes limiting the search to content at a particular selected site. Users may also monitor sites for new content. During formal search on the Web, we expect user to employ search engines that cover the Web comprehensively, and that provide a powerful set of search features to focus retrieval. In order not to miss any important information, there is a willingness to spend more time in the search. Formal search may also be supported by monitoring activity (e.g. e-mail alerts and subscriptions).

Research methods

Participants

A total of 24 participants from 20 organizations took part in the study. Participants' organizations included telecommunications and IT firms, healthcare organizations, small and medium Web design and development businesses, community organizations, government agencies, and public libraries. Most of the participants would be generally regarded as technically proficient Web users whose daily work involved a heavy use of the Web. In most cases, participants were connected to the Internet through LAN, DSL or cable modems, and used relatively high-powered computers. The primary vehicle for publicizing the study was the Internet, through announcements on IT mailing lists (particularly listservs of women in IT associations), and Web sites (such as

those of women in IT associations).

Participants were also recruited using print flyers that were distributed at meetings of a local women's Internet group (Toronto WebGrrls, renamed Digital Eve), and a media design lecture held at the University of Toronto. Notices were posted on bulletin boards in various buildings in downtown Toronto that rented space to IT companies.

The study focuses on women in IT professions for two reasons. First, women form an increasingly important and numerous segment in the IT skilled labor force as well as in the larger community of proficient Web users. Second, most participants of Web studies of this kind are predominantly male. Most studies sampled students and faculty in computer science departments. Research shows that women comprise less than 25 percent of students enrolled in computer science and technology programs and women continue to be under-represented in the IT sector (Spender, 1995; Catalyst 2000a, b; Margolis and Fisher, 2002; Chow, 2002).

Data collection

Two methods of data collection were employed: a WebTracker application that recorded Web browser actions; and personal interviews with participants. History files and bookmark files were also collected.

Web tracking

To observe participants' actual information seeking behavior on the Web in an unobtrusive manner, custom-developed WebTracker software was installed on their computers to record two weeks of continuous Web activity. A ten-day recording period was deemed sufficient since IT workers make heavy use of the Web on a daily basis. WebTracker is a Windows application that logs a Web user's browser actions during a session of Web use (Figure 1).

WebTracker runs in the background and gathers menu, button, and keyboard choices from the browser, including "Open URL or File," "Reload," "Back," "Forward," "Add to Bookmarks," "Go to Bookmark," "Print," and "Copy" (Table IV). As URL calls and menu selections were date-time stamped, we were able to reconstruct sequences of browser actions.

Post-WebTracker interview

After the tracking logs had been collected, each participant was interviewed. During interviews, the WebTracker log files and bookmark file were reviewed with the participant. The researcher presented the participant with printouts of log files and discussed the context of the information seeking episode, sites visited, and Web browser actions. The interview included these questions:

- (1) Can you briefly describe the context behind your use of Web-based information sources?
- (2) By Web-based information sources, I am referring to Web sites and communication sources available on the Web (Web-based chatrooms and bulletin boards), as well as communication sources accessed through a Web browser, such as newsgroups, and listservs (mailing lists).
- (3) Why did you visit the Web-based information sources shown in the tracking log?
- (4) Did you find the information you were looking for?
- (5) Did you run into difficulties looking for this information?
- (6) How did you use the information once it was found?
- (7) Was this an episode of browsing or searching?
- (8) If browsing: which of the following best fits your description of this episode?
 - "I was broadly browsing a variety of Web sites/listservs/chatrooms/newsgroups/boards, taking advantage of what was easily accessible."
 - "I was browsing the same Web site/listserv/chatroom/newsgroup/boards I usually visit for this type of information."
- (9) If searching: which of the following best fits your description of this episode?
 - "I was searching for information in an unstructured way that was 'good enough' to help me solve a problem or move me along."
 - "I was searching for information in a structured way to help me make a decision or set a course of action."

Figure 1 WebTracker application screen with privacy/securing notice

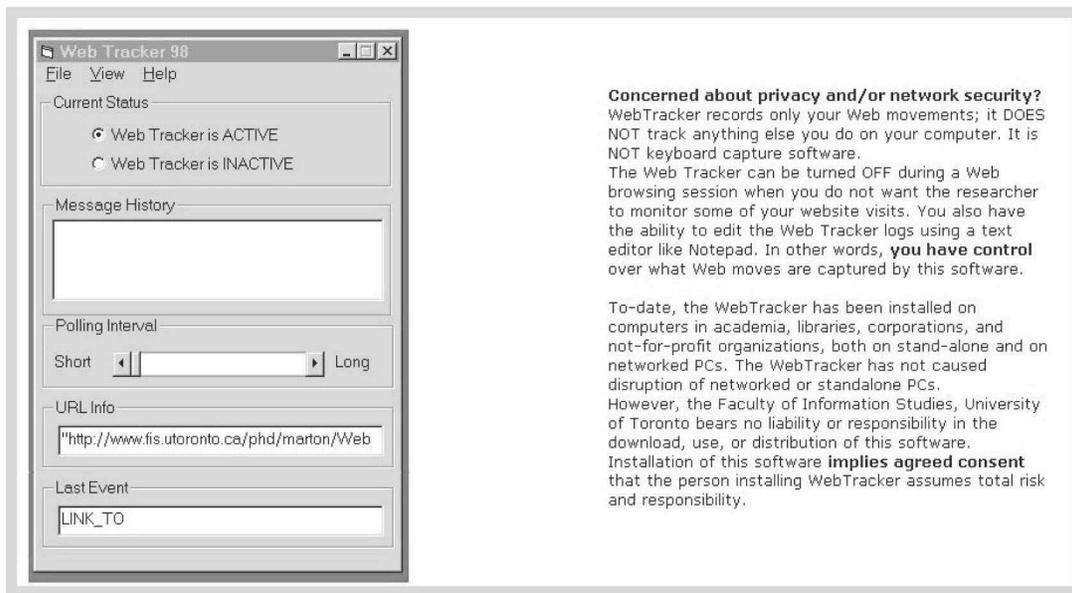


Table IV Example WebTracker log entries

STARTUP	8/25/00 9:36	
LINK_TO	8/25/00 9:51	http://intra.mbs.gov.on.ca/ , "MBS INTRANET", "master"
LINK_TO	8/25/00 9:52	http://jobs.gc.ca/home_e.shtml , "PSC – Recruitment Menu", "master"
key copy	8/25/00 9:52	http://jobs.gc.ca/home_e.shtml , "PSC – Recruitment Menu", "master"
LINK_TO	8/25/00 10:36	http://intra.gov.on.ca/ , "Welcome to the Ontario Public Service Intranet", "August"
LINK_TO	8/25/00 10:37	http://intra.cpb.gov.on.ca/ , "Corporate Policy Branch Intranet", "August"
LINK_TO	8/25/00 11:11	http://www.interland.com/hosting/dedicated/windows/default.asp , "Windows 2000 Plans and Pricing – Interland Web Solutions", "August"
button print	8/25/00 11:36	http://www.interland.com/hosting/dedicated/windows/default.asp , "Windows 2000 Plans and Pricing – Interland Web Solutions", "August"
LINK_TO	8/25/00 11:37	http://intra.cpb.gov.on.ca/ , "Corporate Policy Branch Intranet", "August"
key forward	8/25/00 12:06	http://intra.cpb.gov.on.ca/html/sites.html , "Corporate Policy Branch Intranet", "August"
LINK_TO	8/25/00 12:06	http://www.gbrownc.on.ca/ , "BorderManager Information Alert", "August"
LINK_TO	8/25/00 12:06	http://www.gbrownc.on.ca/sitemap.html , "George Brown College: Site Map", "August"
LINK_TO	8/25/00 12:07	http://www.gbrownc.on.ca/News/faq.html#reg , "Guide to Students' most frequently asked questions", "August"
button back	8/25/00 12:08	http://www.gbrownc.on.ca/News/faq.html#reg , "Guide to Students' most frequently asked questions", "August"
LINK_TO	8/25/00 12:08	http://www.gbrownc.on.ca/News/faq.html , "Guide to Students' most frequently asked questions", "August"
menu forward	8/25/00 14:39	http://www.gbrownc.on.ca/News/faq.html , "Guide to Students' most frequently asked questions", "August"
LINK_TO	8/25/00 14:39	http://intra.cpb.gov.on.ca/ , "Corporate Policy Branch Intranet", "August"
button reload	8/25/00 14:39	http://intra.cpb.gov.on.ca/html/sites.html , "Corporate Policy Branch Intranet", "August"
LINK_TO	8/25/00 14:40	http://intra.managersmap.gov.on.ca/scripts/index.asp , "CSB Manager's Map Log In", "August"

Additionally, participants were asked to recount incidents where they found information on the Web that was valuable and led them to make a significant decision or change the way they did things. While participants were often unable to recall specific incidents, their elaborations of the WebTracker log data revealed many significant episodes of information seeking that were consequential. For the most part, they found that information on the Web verified or added to information they already possessed from other sources. Participants frequently commented that the Web was the only information source that could provide them with current information on software and IT developments.

Data analysis

Data analysis proceeded in two stages. First, significant episodes of Web-based information seeking were identified from the personal interview transcripts as well as the WebTracker logs. Each episode was analyzed according to its information need, amount of effort, number of Web sources consulted, and information use. Using the framework developed earlier, an episode would be categorized, where applicable, as one of the four modes of scanning (undirected/conditioned viewing; informal/formal searching). WebTracker logs were also examined to identify additional viewing or searching episodes. Two criteria were used to select episodes: the episode consumed a substantial amount of time and effort; or the episode was a frequently repeated activity.

For each episode identified in the first stage, corresponding sections of the WebTracker log were examined to determine the browser-based actions that best characterized each episode. Using the framework developed earlier, browser moves were categorized, where applicable, as starting, chaining, browsing, differentiating, monitoring, or extracting.

Findings

A total of 80 episodes of “significant” information seeking were identified and categorized using the framework developed earlier. The majority of the episodes were classified as conditioned viewing (31) and

informal search (21). A smaller number of episodes were undirected viewing (14) and formal search (14).

Of the 80 episodes, 14 were categorized as undirected viewing. The most common example of undirected viewing consisted of visits to general news Web sites such as those of my.yahoo.com, my.netscape.com, and major newspapers (Table V). Some participants have customized these pages as their start-up pages. General news sites acted as gateways to information covering many different subject areas, and provided an efficient way of surveying current developments without a specific information need in mind.

In the undirected viewing episodes, WebTracker data indicated that the most frequently occurring moves were starting and chaining. Thus, participants began at favorite starting pages (news or portal sites) and followed interesting links.

A total of 31 episodes were categorized as conditioned viewing, forming the largest group. The most common examples were regular return visits to bookmarked sites, and browsing from a selected page that contained links to sites of interest (Table VI). Thus, participants regularly visited the Web sites of software and hardware vendors, competitors, Web development resources, and job sites. Several participants subscribed to newsletters and discussion groups, and returned periodically to look at these resources.

In the conditioned viewing episodes, the most frequently occurring moves were differentiating, browsing, and monitoring. Thus, participants selected a bookmarked page/site, or entered the URL of a site they had committed to memory (differentiating). These sites/pages were then examined to locate new content of interest (browsing). The most important characteristic of conditioned viewing was that participants regularly or periodically returned to their selected or differentiated sites/pages to check for new information (monitoring).

A total of 21 episodes were categorized as informal search. The most common examples of informal search were when participants made use of specific query terms such as names of companies, products or technologies to perform simple searches on easily accessible search engines (Table VII). The use of the Google

Table V Undirected viewing examples

Information seeking in the undirected viewing mode	Browser moves: starting, chaining
Kept up to date by scanning ca.my.yahoo.com page daily	Took one afternoon to set up my.yahoo page as start-up page. Clicked on one link per day
Kept up to date with news in torontostar.com about Toronto	Scanned and followed links on site to stories of interest
Kept up to date with daily news in my.netscape.com	Had long used personalized my.netscape.com as gateway
Kept up to date with news in my.yahoo: "go there periodically throughout the day . . . just look at news and things"	Browsed my.yahoo.com Followed links in dailynews.yahoo.com
Kept up to date with news sites canoe.ca, <i>Globe and Mail</i> , <i>Toronto Star</i> . Read headlines	Placed bookmarked sites onto "a Web page that I did long time ago" of frequently used links

Table VI Conditioned viewing examples

Information seeking in the conditioned viewing mode	Browser moves: differentiating, browsing, monitoring
Browsed and looked up postings within discussion group on framemaker in egroups.com	Subscribed to discussion group. Looked for Web site mentioned in discussion group posting
Used alerts.web.ca to view alerts on women's issues	Browsed in alerts.web.ca Followed links
Looked at breaking news in <i>PC World</i> . Looked for items of interest (e.g. news about Oracle)	E-mail subscription. Browsed links in pcworld.com/news/ articles. Followed links to three articles at different times, days
Read up on topics of interest (e.g. WML, wireless, Web design)	Browsed daily e-mail subscriptions to WebMonkey, HotWired, ZDNet, WebReference
Looked at competitor Web sites: "Keep an eye on the competition"	Visited sites of competitors periodically

Table VII Informal search examples

Information seeking in the informal search mode	Browser moves: differentiating, extracting/local search
Searched for new director of research funding agency	Simple Google name search. Also searched for books authored
Searched for info on GIS software tools, and what others are using	Searched Google using simple text string
Searched for ISO 9000 and ACM Professional Chapter Toolkit	ACM site search using local search engine
Looked for groupware product info, led to vendor site Preparing article for newsletter. Looked for Web sites	Simple Google keyword search. Linked to product pages Looked in Association site. Found page on "Medspeak"

search engine was common. There were several examples of selecting search engines that were specific to the Web site (e.g. a search engine maintained by an organization that indexed only its own Web pages). Thus, participants used the search engine on the Web site of ACM, Compaq, and IBM to do local searches.

In the informal search episodes, the most frequently observed moves were differentiating and localized extracting. Thus, participants went directly to selected sites where they expected to find the needed information

(differentiating). Searching at these sites would make use of local search engines that only retrieved information from those sites (localized extracting). Some participants returned to specific sites to repeat their informal searches (monitoring).

In total, 14 episodes were categorized as formal search. Here, participants were intending to use the information formally (e.g. to write reports, to make decisions), and were prepared to invest effort in a more thorough search. A handful of search episodes had an

elapsed time of more than two hours, and a couple were spread over two days (Table VIII).

In the formal search episodes, the most frequently occurring move was extracting. Here, extracting was more extended and extensive than the localized extracting of informal searching. Thus, participants systematically worked through a number of search engines or meta search engines so as to find (all) important information about a topic or item. Formal searches were often accompanied by additional information extracting activities such as printing and saving pages, downloading PDF documents and PowerPoint slides.

Examining the distribution of information seeking episodes by participants' organizations, the data seem to suggest that participants from the large IT or telecommunication firms in this study carried out a greater number of formal searches. Conditioned viewing and informal search were activities common to all types of participant organizations. There were several examples of conditioned viewing among the smaller organizations (community organizations, Web developers, etc.) that involved subscribing to electronic newsletters or listservs, and subsequently clicking on interesting links in e-mailed alerts or in discussion forums. Participants who were employees of a large telecommunication firm also displayed significant use of the corporate intranet.

Figure 2 shows the distribution of information seeking episodes found in this study. Episodes were categorized according to their information needs, seeking, and use. It was possible to categorize all episodes into one

of the four modes of viewing and searching. Moreover, each mode could be characterized by frequently observed browser actions, as predicted by the model developed earlier (note, however, comments on monitoring and extracting later in this section).

Discussion

The study analyzed 80 episodes of information seeking on the Web by 24 women in IT professions over a period of two weeks. Four modes of viewing and searching were observed - undirected viewing, conditioned viewing, informal search, and formal search. For each of these modes, the study highlighted some distinctive patterns of information seeking:

- (1) In undirected viewing, users were approaching the Web as an efficient, one-stop location to scan information from a wide variety of sources in order to keep up to date with news and commentary in many broad areas.
- (2) In conditioned viewing, users followed topics of interest in a two-step procedure. They:
 - bookmarked or remembered sources that were perceived to be valuable in tracking those topics (differentiating); and
 - periodically returned to these sources in order to view new content and thus track developments in those areas (monitoring).
- (3) In informal search, users conducted simple, quick-and-easy searches using general as well as site-specific search engines (e.g.

Table VIII Formal search examples

Information seeking in the formal search mode	Browser moves: extracting
Searched for info on "how to hold meeting" to organize and facilitate a meeting	Extended, multiple searches (food, themes, contests). Printed pages. Elapsed time: 2 hours
Searched for telephone numbers in province needed to test new feature on the provincial telephone switch	Searched Canada 411 for province phone numbers of several businesses. Elapsed time: 75 minutes.
Searched for documents on Cable versus DSL evaluation. Found market research reports from IDC, Forrester	Searched internal host, back/forward moves. Printed several PDF documents. Elapsed time: 3+ hours
Collected information on enterprise security software for project, to write report, make recommendation	Searched internal search engine as well as that on vendor site. To "quote chapter and verse." Over two days
Searched for Internet filtering/blocking software to be installed on library public terminals	Asked two listservs for filtering software. Searched several sites, listservs. Visited product sites

Figure 2 Information seeking modes and browser moves (2002)

	Starting	Chaining	Browsing	Differentiating	Monitoring	Extracting
Undirected Viewing	14					
Conditioned Viewing		31				
Informal Searching				21		
Formal Searching						14

Google, and the ACM digital library search engine). Search queries were constructed simply, using keywords as search terms. Advanced search features such as Boolean operators or word truncation were rarely used.

- (4) In formal search, users were driven by the specific requirements of the task or problem that they had to deal with. Formal search was different from informal search in the use of multiple search engines, and in the greater time that was spent in iteratively exploring search results. However, there was no major qualitative difference in the way that search queries were constructed (which were still based on the use of simple keywords).

In this study, the largest category of information seeking episodes was conditioned viewing (31 instances out of a total of 80). Differentiating and monitoring moves were frequently observed in the conditioned viewing episodes. Thus, participants were able to recognize topics of special interest (differentiating), and they returned periodically to selected sites for new information (monitoring). Participants selected a bookmark or entered a URL to revisit a site; or they would subscribe to mailing lists, or e-mail newsletters about new content on interested sites. Contrary to the model's

predictions, there were no instances of monitoring during formal searching episodes.

The phenomenon of returning to familiar Web sites has been noted in a number of recent field studies. In their analysis of client-side log data, Cockburn and McKenzie (2000, p. 921) observed that users were revisiting Web pages more frequently: "More surprisingly, the revisitation rate has increased from approximately 60 to 81 percent; on average, users have previously seen four out of five pages that they visit." Cothey (2002) noted a trend among more experienced Web users: "... as Web users become more experienced they rely less on formal querying (typically using search engine) to obtain their Web-based information. However, as users become more experienced, so their Web usage was more sporadic, which suggests possible greater selectivity. It appears that each user may inhabit an individual niche of Web hosts that becomes more distinctive as the user becomes more experienced" (Cothey, 2002, p. 77). Cothey (2002) suggests that users adopt a more passive or browsing approach to Web information searching as they gained experience.

In this study, most episodes of informal and formal searching employed basic searching strategies. For the most part, search formulations were relatively simple, with

advanced features such as Boolean operators, and word truncation or proximity operators rarely utilized. This was the case even when participants appeared to be working in the formal search mode. There were no instances of participants accessing search engine help instruction pages to improve their searches. The phenomenon of “simple searching” has been observed in a number of recent studies. For example, Jansen *et al.* (2000) analyzed the transaction log of over 50,000 queries by 18,000 users of the Excite search engine, and found that most queries were short (a query contained 2.21 terms on average, with two in three containing one or two terms); and that Boolean and more advanced operators were seldom used (one in 18 users used any Boolean capabilities, and only one in 190 queries used nested logic).

Women in IT professions exemplify a gendered self and a gendered situation (Howard and Hollander, 1997). The gendered self is manifested in terms of work role. In this study population, women in the IT sector occupy positions that primarily entail communications and/or design; there is a lack of programmers and senior managers. The gendered situation of IT information seeking on the Web is evident in several steps of the information seeking process and is to some extent entwined with the gendered work role. Information needs include seeking information on women-specific topics, such as women’s health, for the two out of three participants who work in health organizations. Another participant designs a women’s Web site, and her information seeking was for the purpose of organizing training and events, and writing reports. For participants who were sole owners of Web design businesses, there was a preference for computer-mediated communication, in particular, the use of women in IT listservs. These served as a source for information about IT, as well as a forum for networking.

In the context of our model of information seeking on the Web, subscribing to women in IT listservs represents differentiating (identifying quality sources), while sending and reading posts from these listservs represents monitoring (keeping track of issues). In some instances, information use is also gendered: there is considerable evidence of sharing of IT

information with family members, workplace colleagues and management, through e-mail and with women in IT professionals, through the use of women in IT listservs.

At a more general level, the study has implications for the design and evaluation of the intranet or other Web-based infrastructure for supporting knowledge work. In designing a Web site, the study suggests that the design needs to facilitate not only the finding or searching for information, but also the browsing or viewing of information. Browsing - as in undirected and conditioned viewing - is the principal way by which organizational participants develop, elaborate, and prioritize their emergent information needs. As organizations thrive in complex and open environments, the information skill to watch and assess emergent trends and phenomena, through scanning and the broadening of peripheral vision, will become increasingly salient.

For evaluating a Web site, the study suggests that an analysis of browser-based actions and events initiated by users as they look for information may be a method for evaluating the usability and usefulness of the site. By elucidating patterns of browser actions or moves - cycles of stepping back and forth between Web pages, climbing up and down a Web site hierarchy; bookmarking; printing; and so on - we can see the preferences of users, their information seeking styles and strategies, as well as the occasions when they fail to locate usable information. This fine-grained feedback based on actual use can be analyzed to enhance the usefulness of a Web site or portal.

At a practical level, we may postulate ways of supporting and enhancing the information seeking of Web users as they engage in these four modes of information seeking (Table IX, adapted from Choo *et al.*, 2000).

Summary

The research presented here developed a new conceptual model for describing information seeking behaviors on the Web by integrating constructs derived from organization science and information science. The conceptual model was then applied empirically in a field study of

Table IX Enhancing Web use

Web use mode	Enhancing Web use
Undirected viewing: starting and chaining	Introduce systems that can search or recommend new starting-sites or "similar-to" sites Encourage people in a group to share bookmarks, Web pages, URLs Design corporate portals to support undirected, broad-based viewing
Conditioned viewing: browsing, differentiating, monitoring	Train users to evaluate and escalate priority or importance of information Make it easy for users to share Web-based information via e-mail or in online forums Introduce users to services that allow them to subscribe to and be notified about new content on Web pages
Informal search: differentiating, monitoring, extracting	Pre-select high quality sources and search engines for quick, informal searches Prepackage good search strategies developed by subject matter experts; allow users to view these strategies and learn from them Educate users on how to evaluate information provenance and quality
Formal search: extracting	Educate users about full range of information sources that should be considered for comprehensive search: print, online, human sources Educate users about when to use: commercial online database services; the library or information resource center; information brokers/professionals Train users on advanced search techniques: narrowing or broadening a search; balancing precision and recall; backward and forward chaining

Web use by women in IT professions. The model proved useful in analyzing the data collected and the analysis yielded interesting insights about browsing and searching for information on the Web.

Thus, in line with the conceptual model, four complementary modes of information seeking were observed: undirected viewing, conditioned viewing, informal search, and formal search. In each mode of viewing or searching, users would adopt distinctive patterns of browser moves: starting, chaining, browsing, differentiating, monitoring, and extracting. Each mode of information seeking on the Web is also distinguished by the type of information needs, information seeking tactics, and the purpose of information use. The information seeking tactics characterizing each mode are revealed by recurrent sequences of browser actions initiated by the information seeker. Overall, the study suggests that a behavioral framework that links modes (strategies and reasons for viewing and searching) and moves (actions used to find and use information) may be helpful in analyzing Web-based information seeking.

The study also suggests that multiple, complementary methods of collecting qualitative and quantitative data may be integrated within a single study to compose a more nuanced portrayal of how individuals seek

and use Web-based information in their natural work settings. The data collection approach adopted was multimodal, embracing both interpretive data elicited directly from the users, and interactive data logged unintrusively by software. Thus, users' personal stories of their episodes of finding and making use of significant information furnished the narrative frameworks in which the detailed data on browser actions and site visits were interpreted. These users' stories sometimes elaborated the thinking processes, emotional states, and situational constraints that contextualized the episodes of encountering information. Our experience in this study indicates that a research design combining qualitative and quantitative data can offer the depth and detail needed to make sense of Web-based information seeking and use.

References

- Aguilar, F.J. (1967), *Scanning the Business Environment*, Macmillan Co., New York, NY.
- Catalyst (2000a), *Women in Information Technology*, InfoBrief, available at: www.catalystwomen.org/press/infobriefs/infombatech.html
- Catalyst (2000b), *MBA Grads in Info Tech: Women and Men in the Information Technology Industry*, Catalyst, New York, NY.

- Catledge, L.D. and Pitkow, J.E. (1995), "Characterizing browsing strategies in the World Wide Web", available at: www.igd.fhg.de/www/www95/papers/80/userpatterns/UserPatterns.Paper4.formatted.html
- Choo, C.W., Detlor, B. and Turnbull, D. (2000), *Web Work: Information Seeking and Knowledge Work on the World Wide Web*, Kluwer Academic Publishers, Dordrecht.
- Chow, J. (2002), "Statistics show fewer women in IT careers. ITworld.com", available at: http://itworld.com/Career/1832/020410itcareers/page_1.html
- Cockburn, A. and McKenzie, B. (2001), "What do Web users do? An empirical analysis of Web use", *International Journal of Human-Computer Studies*, Vol. 54 No. 6, pp. 903-22.
- Cothey, V. (2002), "A longitudinal study of World Wide Web users' information-searching behavior", *Journal of American Society for Information Science & Technology*, Vol. 53 No. 2, pp. 67-78.
- Ellis, D. (1989), "A behavioural model for information retrieval system design", *Journal of Information Science*, Vol. 15 No. 4/5, pp. 237-47.
- Ellis, D. and Haugan, M. (1997), "Modelling the information seeking patterns of engineers and research scientists in an industrial environment", *Journal of Documentation*, Vol. 53 No. 4, pp. 384-403.
- Ellis, D., Cox, D. and Hall, K. (1993), "A comparison of the information seeking patterns of researchers in the physical and social sciences", *Journal of Documentation*, Vol. 49 No. 4, pp. 356-69.
- Hölscher, C. and Strube, G. (2000), "Web search behavior of Internet experts and newbies", *Computer Networks*, Vol. 33 No. 1-6, pp. 337-46.
- Howard, J.A. and Hollander, J.A. (1997), *Gendered Situations, Gendered Selves: A Gender Lens on Social Psychology*, Sage Publications, Thousand Oaks, CA.
- Huberman, B.A., Pirolli, P.L., Pitkow, J.E. and Lukose, R.M. (1998), "Strong regularities in World Wide Web surfing", *Science*, Vol. 280 No. 5360.
- Jansen, B.J., Spink, A. and Saracevic, T. (2000), "Real life, real users, and real needs: a study and analysis of user queries on the Web", *Information Processing & Management*, Vol. 36 No. 2, pp. 207-28.
- Kim, K.-S. and Allen, B. (2002), "Cognitive and task influences on Web searching behavior", *Journal of American Society for Information Science & Technology*, Vol. 53 No. 2, pp. 109-19.
- Marchionini, G.M. (1995), *Information Seeking in Electronic Environments*, Cambridge University Press, Cambridge.
- Margolis, J. and Fisher, A. (2002), *Unlocking the Clubhouse: Women in Computing*, MIT Press, Cambridge, MA.
- Spender, D. (1995), *Nattering on the Net: Women, Power and Cyberspace*, Garamond Press, Toronto.
- Tauscher, L. and Greenberg, S. (1997a), "How people revisit Web pages: empirical findings and implications for the design of history systems", *International Journal of Human-Computer Studies*, Vol. 47, pp. 97-137.
- Tauscher, L. and Greenberg, S. (1997b), "Revisitation patterns in World Wide Web navigation", in Pemberton, S. (Ed.), *CHI 97 Human Factors in Computing Systems*, CHI 97, Atlanta, GA, pp. 399-406.
- Weick, K.E. and Daft, R.L. (1983), "The effectiveness of interpretation systems", in Cameron, K.S. and Whetten, D.A. (Eds), *Organizational Effectiveness: A Comparison of Multiple Models*, Academic Press, New York, NY, pp. 71-93.
- Wilson, T.D. (1997), "Information behaviour: an interdisciplinary perspective", *Information Processing & Management*, Vol. 33 No. 4, pp. 551-72.

Further reading

- Choo, C.W., Detlor, B. and Turnbull, D. (2000), "Working the Web: an empirical model of Web use", presented at the 33rd Hawaii International Conference on Systems Science, Maui, HI.
- Daft, R.L. and Weick, K.E. (1984), "Toward a model of organizations as interpretation systems", *Academy of Management Review*, Vol. 9 No. 2, pp. 284-95.
- Furber, R. (1998), *Does Jane Compute? Preserving Our Daughters' Place in the Cyber Revolution*, Warner Books, New York, NY.