Collaborating Alone and Together: Investigating Persistent and Multi-User Web Search Activities

Meredith Ringel Morris
Microsoft Research
One Microsoft Way
Redmond, WA, USA
merrie@microsoft.com

ABSTRACT

Today's web browsers provide limited support for rich information seeking and sharing scenarios. A survey we conducted of 204 knowledge workers at a large technology company has revealed that a large proportion of users engage in searches spanning multiple sessions and machines, and that such searches may include collaborative activities. We present the results of the survey, and then review the implications of these findings for designing new web search interfaces that provide tools for resumption of state and for sharing. Guided by the survey results, we propose a set of key collaboration features, including the importance of providing support for *persistence*, *awareness*, and *division of labor*.

Categories and Subject Descriptors

H.3.3 [**Information Systems**]: Information Search and Retrieval – *search process.*
H.5.3 [**Information Interfaces and Presentation**]: Group and Organization Interfaces - *computer-supported cooperative work.*

General Terms

Human Factors.

Keywords

Web search, exploratory search, informational search, persistent search, computer-supported cooperative work.

1. INTRODUCTION

In order to create a more detailed picture of users' habits and needs when tackling complex exploratory searches, we conducted a survey of knowledge workers at a large corporation. Our survey findings revealed that it is surprisingly common for web search tasks to span time, devices, and users. The majority of people surveyed reported conducting web searches that spanned multiple sessions separated by time gaps ranging from hours and days to weeks and months. A large number also reported conducting web searches that began on one computer and were resumed on another machine. And nearly everyone in our survey population indicated that they had collaborated with others, sharing either the process or products of a web search task. None of these three scenarios (resuming a search after a time delay, continuing a search across multiple devices, or collaborating with others to search) is well-supported by standard user interfaces for web search provided by search engine services and client-side browsers.

SIGIR'07, July 23–27, 2007, Amsterdam, Netherlands.

The first two scenarios, searching across time and across devices, can be considered as a special case of the third. Resuming a search after a time delay can be viewed as a case of asynchronous collaboration with oneself, and resuming a search after switching devices can be viewed as a case of remote collaboration. Thus, it is reasonable to envision a single search interface to support collaboration that could simultaneously assist with all three of these scenarios.

In this paper, we present the findings of our survey on collaborative web search habits and needs, which includes data about both "collaborating alone" (*i.e.,* continuing a search across time or devices) and "collaborating together" (*i.e.,* a search task that involves multiple people). We then synthesize recommendations for enhancing web search interfaces. These recommendations, based on the trends identified through our survey, support both collaboration with others as well as collaboration with oneself across time or devices. We conclude with a discussion of related work, including studies of search practices ([1][3][14][16][20]), of cooperative information retrieval ([5][7][10][11][12][13][19]), and of re-finding information ([2][9][17][18]).

2. SURVEY

In order to understand current practices and needs regarding collaboration on web search tasks, we created a twenty-five question survey, eliciting both multiple-choice and free-form responses regarding respondents' web search tactics for complex, exploratory searches. These questions focused particularly on searches that span several sessions and/or computers, and on searches which involved multiple stakeholders. This survey was distributed online in November 2006 to 740 workers at a large technology company. 204 respondents completed the survey (response rate = 27.6%). Entry into a prize drawing served as incentive for completion of the questionnaire.

2.1 Demographics

80.4% of respondents were male. Respondents' ages ranged from 21-61 years old, with a median age of 36 years. Respondents had various roles within the company: 38.2% were researchers, 21.6% were software developers, 16.9% were program managers, and the remaining 23.3% held a mixture of jobs, including managers, administrative assistants, interns, and attorneys.

Respondents all self-identified as relatively sophisticated web searchers. When asked to rate their ability to successfully find

information using a web search engine as either "novice," "average," or "expert," 73.5% chose "expert," and 26.5% chose "average." No respondents classified themselves as novice users of search engines. As an additional indication of the search experience/expertise of participants, we asked how often they used a web search engine. 91.7% responded "several times per day," 7.8% responded "once a day," and only 1 respondent indicated that he/she searched the web less frequently than once per day.

We targeted our survey to this demographic as they tend to be technology leaders who might push the boundaries of envisioned usage scenarios for search tools. Although the specific numbers we report in this paper may not generalize beyond this population, we believe that many of the behaviors and tasks types described are more generally applicable; for instance, many respondents mentioned collaborating on search tasks with non-work colleagues (*e.g.*, spouses, parents, children, friends).

2.2 Collaborating Alone: Persistent Search

In this section, we present survey findings about search tasks that extend across multiple sessions or that use multiple computers. Although these are single-user search tasks, it is fruitful to consider them as instances of either asynchronous or remote "collaboration" where collaborations are with oneself over time. Such a perspective reveals the potential for a single user interface design to address the needs of not only the scenarios described in this section, but also of the "collaborating together" scenarios detailed in section 2.3.

2.2.1 Prevalence

Exploratory search tasks often span multiple sessions and/or computers. 37.3% of respondents answered "Yes" in response to the question *"Have you ever used more than one computer to complete a single web information-seeking task? (i.e., some combination of office computer, laptop, home computer, PDA, etc.)".* Our survey also asked, *"Have you ever had a web information-seeking task that has lasted for longer than a single session? (i.e., a task which you put aside and resumed at a later time)."* 83.3% of the 204 respondents answered "Yes" to this question.

To further probe this issue, we asked a follow-on question to the people who responded that they had, indeed, conducted a multi-session web search: *"Which of the following best describes the longest time gap you have experienced between leaving off and resuming a web information-seeking task?"* The results in Table 1 show that time gaps before resuming a suspended search task can be quite lengthy, with 74.7% of respondents indicating they have experienced gaps of at least one day before resuming a search task, and 24.1% having resumed a search task after a postponement of a week or longer.

Table 1. The 170 respondents who had resumed a web search after a gap in time reported the longest gap they recalled between suspension and resumption.

Longest Time Before Search Resumption	**Respondents**
Minutes	1.2%
Hours	24.1%
Days	50.6%
Weeks	12.9%
Months	10.0%
Years	1.2%

2.2.2 Strategies for Task Resumption

We also asked the people who had experienced a multi-session and/or multi-computer search, *"What strategies do you use to refamiliarize yourself with the context (goals, current state, etc) of your web information-seeking task after [a time gap/switching computers]?"*. These questions allowed for free-form responses. We classified similar responses according to the strategies used to assist task resumption.

The main strategies that emerged can be described along two axes: *initiative* (whether the user was *active* or *passive*) and *stage* (whether the strategy applies to the *storage* or *retrieval* of the search context). For example, an active approach to storage attempts to preserve the state, strategies, and/or results of a search task for later use, while a passive approach to storage involves no explicit user actions to save information during a search session. An active approach to retrieval seeks out mechanisms that have captured data (either explicitly or implicitly) that may help in re-establishing a search context, while a passive retrieval strategy uses only information that is automatically revealed by basic browser or search engine interfaces. We describe strategies by indicating the initiative type for each of those stages.

Active Storage/Active Retrieval strategies reported in our survey included saving the contents of a website (either into a local file or by printing), bookmarking a website (either locally in the browser or via an online bookmarking utility), and note-taking (either by hand, via a typed document, self-email, or instant messages). Examples descriptions of these strategies included:

- "email search results to self to remember the queries that got the right results"[1]
- "save aggregation sites to favorites"
- "If I realize that it will take a longer time, I create a folder under favorites and store the most relevant sites I located."
- "notes on paper or in text document/Excel spreadsheet"
- "Delicious [http://del.icio.us, a tagged bookmarks site]"

Passive Storage/Passive Retrieval strategies reported by our survey population included relying on memory when resuming a search (and possibly starting over and repeating previously issued queries), leaving the browser open on the machine when suspending a task so that its state was preserved upon resumption, relying on the browser/search engine's autocomplete functionality to serve as a memory aid when typing queries, and relying on the browser's coloring of visited links to prompt memory about previously viewed web sites. Example descriptions of these strategies included:

- "Usually start all over again"
- "I tried to remember which search query keywords had produced positive results and/or where I had left off. Basically, I re-traced my steps."
- "Don't close my old browsers that were searching!"
- "I used the auto complete feature in the search box"
- "[I use] link visited color."

[1]Spelling errors have been fixed when reporting survey responses.

Examples of Passive Storage/Active Retrieval strategies include opening up the browser's history of previously-visited sites in order to find information from a previously executed search task, or visiting a search-engine-specific query history tracker in order to re-find a useful query or avoid repeating previously-issued queries. In the cross-machine scenario, transferring data to the new machine was another approach. For example:

- "[I use] IE page history to reconstruct queries."
- "[I] manually copy the search string from one device to the other."
- "Saving of Search Queries in the account itself like Google search history."

Table 2 shows the percentage of people in our survey population who reported employing each of the above strategies, in both the across-time and across-computer scenarios (note that the numbers total to more than 100%, as some users listed several of these strategies). Note that no respondents described using an Active Storage/Passive Retrieval strategy, which is not surprising given that such a combination represents wasted storage effort.

Table 2. Percent of respondents who mentioned using the following strategies when resuming a search after a gap in time (n=170) or on a different computer (n=76).

	Strategy	Prevalence when "collaborating alone" across time	Prevalence when "collaborating alone" across computers
Active Storage/ Active Retrieval	Browser bookmarks	25.9%	N/A
	Online bookmarks/tagging	1.8%	9.2%
	Save web page to file	5.3%	1.3%
	Print web page	0.6%	1.3%
	Written or typed notes	30.0%	13.2%
	E-mail to self	2.4%	23.7%
	IM to self	0.0%	3.9%
Passive Storage/ Active Retrieval	Browser history	7.6%	N/A
	Online query history	1.2%	2.6%
	Manually copy data from display	N/A	5.3%
Passive Storage/ Passive Retrieval	Leave browser open	14.1%	N/A
	Autocomplete	2.4%	N/A
	Link coloring	2.4%	N/A
	Memory	35.9%	35.5%

2.3 Collaborating Together: Multi-User Search

Web search is generally considered to be a solo activity. All major search engines (*e.g.,* Google, Yahoo!, Windows Live Search) and web browsers (*e.g.,* Internet Explorer, Firefox) are designed for use by a single person working alone. However, many tasks in both professional and casual settings can benefit from the ability to jointly search the web with others. Our intuition was that such situations might be commonplace. Thus, we included questions on our survey to determine whether people need and/or want to collaborate when searching the web, and, if they do, what strategies they employ to collaborate given that this ability is not supported in an explicit manner by current search interfaces.

2.3.1 Prevalence

When asked, *"Have you ever cooperated with other people to search the web?"*, 53.4% of our respondents answered "Yes." However, this figure likely underestimates the proportion of our respondents who cooperatively search, as a subsequent question naming specific cooperative search activities (*i.e.,* techniques to approximate a cooperative search experience in the absence of explicit UI support) received "yes" answers from even more people. Only 2.9% of respondents did not report engaging in any of the set of sample collaborative search activities listed in Table 3. Additionally, of the 46.6% of people who said they had not cooperated with others to search the web, 10.5% indicated that they have *"needed/wanted to cooperate with other people to search the web and been unable to effectively do so."*

Table 3. Responses by the 204 respondents about engaging in multi-user search activities associated with varying levels of collaboration on the process or products of web search.

Collaborative Activity	Respondents
Watched over someone's shoulder as he/she searched the web, and suggested alternate query terms.	87.7%
E-mailed someone links to share the results of a web search.	86.3%
Showed a personal display to other people to share the results of a web search.	85.3%
E-mailed someone a textual summary to share the results of a web search.	60.3%
Called someone on the phone to tell them about the results of a web search.	49.0%
Printed web pages on paper to share the results of a web search.	41.2%
Created a document (other than a web page or email) to share the results of a web search.	34.3%
Instant-messaged other people to coordinate real-time web information-seeking.	30.4%
Used a large form-factor or projected display to share the results of a web search.	24.5%
Performed a web search on a projected or large form-factor display during a group meeting.	23.5%
Divided up responsibilities for a search task among several people, and then shared the results.	18.1%
Created or posted to a web page to share the results of a web search.	15.2%

2.3.2 Frequency

Of the 109 people who self-identified as having cooperatively searched the web (not including the additional respondents who indicated performing the actions listed in Table 3), we asked about the frequency with which they engaged in joint web search tasks. This data is summarized in Table 4. Over a quarter of respondents cooperated on a weekly basis, and over three-quarters on at least a monthly basis, which is surprisingly frequent given that web search technologies do not explicitly support this type of cooperative activity.

Table 4. Frequencies of collaborative behavior reported by 109 respondents who indicated they had cooperated with others on web search.

Frequency of collaborating on web search	Respondents
Daily	0.9%
Weekly	25.7%
Monthly	48.6%

Yearly	24.8%

2.3.3 Configurations

We also questioned these 109 self-identified "cooperators" about the configuration of the group involved in cooperatively searching the web. We asked whether they were always located in the same room as the people with whom they were cooperating. 22% indicated they were "always" co-located when cooperatively searching, 11.9% indicated they were "never" in the same room with their cooperators, and the remaining 66.1% indicated that they "sometimes" were co-located with their search partners and sometimes cooperated remotely.

We also asked these same respondents *"What is the typical size of the group of people you cooperatively search with (including yourself)?"*. 80.7% reported a group size of two people, while 19.3% reported a three or four person group, with no respondents reporting a larger group size.

2.3.4 Methods

To the 109 self-identified "cooperators," we also asked a free-form response question: *"Please describe the method(s) by which you cooperate in order to search the web (i.e., how is work divided among participants, what devices are used, how are results shared, etc.)"*.

Respondents reported sharing both the process (*e.g.*, search terms, search sites) and products (*e.g.*, useful links, facts found within sites) of a search with remote collaborators. They indicated three primary means of accomplishing this sharing: email, IM, and phone calls. 20.2% of people reported using email for this purpose, 19.3% used IM, and 15.6% spoke on the phone. No respondents mentioned using special tools or sites for collaboration, such as shared bookmarking or tagging sites.

One aspect of the search process on which many respondents mentioned cooperating was the generation and refinement of query keywords. 22.0% of respondents explicitly mentioned that they cooperate by brainstorming or suggesting keywords to others. Example descriptions included:

- "[We] brainstorm in a group for the best keyword."
- "[We make] joint proposals on query refinement."

When all participants in a joint search task had their own device available to use, the group tended to use one of two search strategies: *divide-and-conquer* or *brute force*.

The *divide-and-conquer* strategies involved explicit coordination and planning, such as assigning different search engines or reference sites to different group members, explicitly dividing up the space of keywords among group members, or explicitly dividing the search task into sub-tasks for each group member to tackle (*e.g.*, planning a vacation divided into finding plane tickets, finding hotels, and finding tourist attractions). 19.3% of respondents mentioned utilizing this type of strategy. For example:

- "[We] use different search engines, search terms."
- "If the search has various aspects we might split them and each of us would do a part of it."

The *brute force* strategies, on the other hand, did not involve any explicit coordination of the search process. Instead, all parties searched separately, possibly duplicating the efforts of other group members, and results were merged afterwards. 24.8% of respondents mentioned using this approach. One respondent described this strategy as:

- "We both search independently using our own individual strategies. Whomever comes up with an interesting result sends it to the other. We usually share results through e-mail or, less commonly, through Instant Messenger."

This class of strategies sometimes takes on aspects of a "race," to see who can find the information fastest. In fact, seven respondents in our survey used vocabulary like "race" or "contest" to describe their multi-user search strategies, giving descriptions such as:

- "I've had Google-races."
- "There is less a division of labor here, than a competition to see who can come up with the most interesting (or entertaining) results."

Co-located users sharing a single device tended to follow a *backseat driver* approach, where one person controlled the mouse and keyboard and the other(s) looked over the "driver's" shoulder, suggesting either keywords to enter or links to examine in more detail. 87.7% of all 204 respondents reported having engaged in this type of behavior. Example descriptions included:

- "If we are in the same room, typically one person 'drives' and the other looks over his/her shoulder and offers suggestions as needed."
- "[We] look at the same screen and one person does the search while the other advises and does 'backseat driving.'"
- "It was more like extreme programming where one person was driving and the other giving suggestions for better results."

2.3.5 Tasks

These same 109 respondents were also asked to *"describe the task (or tasks) for which you have executed a cooperative web search."* Although the responses were free-form, there were quite a few common themes in the tasks people indicated cooperating on, spanning both business and personal topics.

Purchasing items online (for personal or business use) was one major task tackled through cooperative web search. In addition to general online shopping tasks, two sub-categories of shopping were specifically mentioned often enough to deserve separate note: making travel reservations and looking for real estate. For instance, "Researching travel info for a group trip, to match budgets & personal tastes (flights/airfares, accommodations, rental cars, restaurants, local attractions/activities)."

Searching for medical information relevant to an ill family member was also identified by several respondents as a cooperative activity. For example, one respondent described cooperating with his spouse on an internet search while "Diagnosing why our baby was sick one evening."

Also, social planning tasks (choosing a restaurant, finding movie schedules, getting directions to an event, preparing for a party) were frequently cited as multi-user web search activities. For instance, one user described cooperatively searching when

"Looking for a good movie to view together, either in-theatre or rental; picking a local restaurant."

Job-related tasks were also mentioned frequently. For our survey population, which contained a large proportion of researchers and software engineers, this meant that finding academic literature for a jointly-authored research paper was a common multi-user search task, as was searching for technical information (APIs or SDKs, or the meaning of mysterious error messages). One respondent noted, "I have searched for balanced cuts in hypergraphs with my coauthors."

Finally, general fact-finding (in support of a discussion or debate, for curiosity, to assist a child with homework, etc.) was also a task that was sometimes carried out through a joint web search. One respondent mentioned that he jointly searched the web "to find more information about a topic that we have just had or started a discussion about."

Table 5 shows how many people reported executing a cooperative web search for each of these task types.

Table 5. The percentage of people (n=109) whose descriptions of the tasks that they cooperatively search the web to accomplish included the following. The numbers total to higher than 100%, since some people reported doing several of these tasks jointly.

Task	% of Respondents
General shopping tasks	25.7%
Real estate	6.4%
Travel	27.5%
Medical information	6.4%
Social planning	12.8%
Literature search	20.2%
Technical information	16.5%
Fact finding	16.5%

2.3.6 Obstacles

Respondents also had the opportunity to answer a free-form question describing occasions when they had wanted to perform a cooperative web search, but were unable to. The following sample responses to that question illustrate the key themes that we saw throughout the survey as to why current search interfaces don't adequately support cooperation:

Theme 1: Desire to parallelize task without unnecessary duplication of effort. "We were trying to do a lit search but we both have different strategies for how we traverse the space. It was difficult to do together (because we wanted to follow different paths) and doing it together was less productive, however, when we did it separately we weren't sure how much redundant information we were gathering."

Theme 2: Difficulty in helping remote collaborators to navigate to the same content for shared context/focus. "We were in separate houses, both searching for hotels online. Difficulty is in comparing hotels. Things like 'go to Google. Type X. Click the third result. See the link on the left that says "more"? Click that. The second hotel is called Y -- how does that look to you?' are inefficient, but eventually get the job done."

Theme 3: Not realizing the need to share the results of a search until after it is finished. "Not realizing I wanted to share the key pages of information gain across the session until after it was complete. Browser history mechanism are too weak to figure out the good stuff in the trail."

Theme 4: Inadequacy of search UIs for teaching search skills, assisting less experienced users. "Helping less computer-savvy users search the web (e.g. my parents). Resolved by doing the search myself and then emailing the links."

3. DESIGN IMPLICATIONS

Systems for computer-supported cooperative work (CSCW) are often classified along two axes: whether they support *remote* or *co-located* users, and whether these users collaborate in a *synchronous* or *asynchronous* fashion. Based on the results of our survey, we have formulated design guidelines for web search interfaces that support remote collaborators, working either synchronously or asynchronously, which we present in section 3.1. Section 3.2 discusses possible extensions of our design guidelines to support co-located collaborative search.

3.1 Implications for Remote Collaboration

3.1.1 Persistence

Storing a search session in a persistent format is a key requirement for facilitating collaboration during the session, revising the search at a later time, or sharing the results of a search with others. Our survey showed that many people currently use ad-hoc methods in order to make some aspects of search sessions persistent, such as taking notes on query terms used or sites found, or using bookmarking tools. Built-in browser support for saving and loading search sessions could centralize many types of persistence in one location. A persistent representation of a search should have several important qualities:

"Topic" model. Each persistent search session should be related to a particular topic or task. This differs from models used by current technologies, such as web browsers' history features (which store links visited chronologically) or search engines' query histories (which show recent query keywords chronologically). As our survey showed, it is common for people to return to a search task after a gap in time; during this interval they may have performed several other search or browsing actions. Purely chronological organizations of search processes (*e.g.*, query terms used) or products (*e.g.*, links found) interleave these different tasks, making it difficult for users to re-find or share information.

Implicit capture. Information relevant to the search session should be captured and saved implicitly, without the user needing to take active steps in order to create the resulting persistent record of the search process and products. As our survey showed, passive strategies for trying to store/retrieve the context of a search task were much more prevalent than active ones. Based on this data, it seems unrealistic to expect users to devote much effort to creating a persistent search state, since the payoff for this work occurs in the future (or benefits others). Even if users were willing to put in effort to explicitly denote which aspects of their search should be saved, another difficulty is that users don't always realize when they begin a search task that they will need to share it with others, or that they will be interrupted, or even that what seems like a simple search initially might grow into a complex exploratory task. Thus, implicit capture also ensures that users will have the information from a search session preserved even if they may not

initially have realized they or their companions might need to revisit it. Of course, manual means of capturing session information could always supplement an implicit system.

Capture both process and products of search. A persistent search session should include not only the final websites (or sub-portions of websites) that contain the results sought, but should also include information such as the keywords and search engine used to find those results (or the path of links followed from a search engine result to the final "useful" site), since this information is helpful to collaborators in understanding both what techniques have already been tried and in understanding how to interpret the authoritativeness or appropriateness of the results. Additionally, pages found "on the way" to the real results are also worth storing, both so that collaborators (or the original user) can avoid duplication of effort by not re-visiting those pages, and so that information found in passing can be re-found later. In addition to storing URLs, so that updated versions of pages can be retrieved if the session is re-accessed at a later point in time, it would also be useful to store cached versions of pages from the time when they were originally added to a session. This feature is motivated by comments from our survey respondents, which indicated that one frustration in trying to coordinate a multi-person search is that, due to algorithmic changes or changes in underlying content, telling a partner to type certain keywords into a search engine doesn't always guarantee they will see exactly the same results as the other person, or that when they find the page the information sought will still be there. Thus, having cached versions of pages available is useful for providing assured common ground to collaborators when desired, or for preserving data that changes when a search is revisited after a time delay.

Summarization. There are a variety of levels of collaboration during a search task. Very closely-coupled collaboration might mean that all involved parties actively participate in both the search and sensemaking [15] processes, while a more loosely-coupled collaboration might mean that one person does most of the searching and only shares the results of what he has found with his collaborators. The detailed data captured and stored in a persistent search session is very useful to all involved parties in the closely-coupled scenario, but could be distracting and irrelevant to someone who only wishes to share in understanding the results of an informational search. The ability to create summary views of a persistent session would be valuable for more lightweight collaboration scenarios.

3.1.2 Awareness

Another key requirement for collaborative search is a means of providing awareness of the activities of collaborators. This type of information is useful in both synchronous and asynchronous settings. Based on our survey data, the following types of awareness information seem especially valuable to include in a collaborative search interface:

Keyword awareness. A very common scenario reported in our survey was that people collaborated to brainstorm keywords or to offer query reformulation suggestions to each other. By making users aware of the keyword syntax and combinations being entered by their partners, several benefits emerge. First, undesired duplication of effort can be avoided, since a user knows what keywords have already been tried. Second, creativity (or even learning) could be prompted – seeing a keyword entered by a partner might make a user think of a new way to rephrase the query or might demonstrate advanced syntax that the user can experiment with.

Visitation awareness. It would also be useful for a collaborative search interface to visualize which websites/search results have already been viewed by oneself or by other members of the group (and also *when* those results were viewed). This is helpful for avoiding undesired duplication of effort, since users can focus on exploring results/sites not yet visited by group members, as well as for implicit gathering of information about the session – filtering the complete persistent session by only showing sites visited by certain group members might be one technique for summarization.

Markup. In addition to simply knowing which sites have been viewed by collaborators, being able to see markup on the pages themselves (*e.g.*, comments, tags, ratings, and/or hiliting sub-sections of the page) is useful for allowing users to understand what their partners thought of the page, and why they had visited it or felt that it was or was not relevant to the task at hand.

Direct communication channel. Providing a channel for real-time communication (in the event that multiple users are searching synchronously), whether over a textual chat or voice channel, would also be a useful feature to integrate into the search environment. These messages should also be persistent, to allow for asynchronous communication. In our survey, users reported using phones, instant messaging, and email to communicate with others and coordinate the search process. Integrating communication directly into the search application reduces the overhead of this type of coordination, as well as making it possible to capture this aspect of the process for later review and searching.

Attribution. It's important that information stored in a search session, such as sites found, keywords used, etc., have attribution information associated with them such that users are aware which group member found each result or suggested each query formulation. This is useful not only for allowing users to make more accurate judgments about utility and trustworthiness of information, but also for admitting the role of ego in some collaborative search scenarios, potentially further motivating application use. Several users in our survey reported that they view collaboration on a search task as a competition or race, and are motivated to try to find the answer more quickly than their teammates. Proper attribution may often be important to this group of users in particular.

3.1.3 Division of labor

Many of our survey respondents described ad-hoc methods to avoid duplication of effort during a search task, such as dividing up the space of potential keywords, search engines, or sub-tasks among different group members. Supporting mechanisms for dividing up and sharing work among participants is important to the success of a UI for multi-user search. Such mechanisms should support the following abilities:

Close and loose coupling. At times, it is useful for all members of a group to view the same content (*i.e.*, have "yoked" views and each see the same thing on their displays), such as when discussing the contents of a particular page, or when an expert searcher is assisting a more novice searcher. However, systems that only allow yoked or master/slave views, such as some of the early collaborative web browser prototypes [6], prevent the ability to avoid duplication of effort and divide labor among group members. Thus, the ability to not always have yoked views, but to

allow users to independently enter keywords and view results irrespective of what their teammates are currently doing, is also valuable.

Divide and conquer. The ability to have a collaborative search application assist in the division of labor among participants could reduce coordination overhead, giving users more time to examine search results rather than plan how to divide the task space. For instance, one group member could enter a keyword, and then choose a "split results" option, which might divide up the search results from the search engine of choice among the different group members (*i.e.*, user 1 sees all the odd-ranked results and user 2 sees all the even-ranked results). Or, choosing a "cross engine" option might send the same query to several search engines, and give back the results from different sources to different users (*i.e.*, user 1 gets all the Google results and user 2 gets all the Windows Live Search results for the same query). Providing access to a variety of search sources (both whole-web search engines, as well as more task-specific providers such as Wikipedia, the ACM Digital Library, medical databases, etc., would be a valuable feature; several respondents to our survey mentioned using a variety of sources for a single search task, or dividing up sources among group members. The ability to capture the history of interaction with all of these sources in one place, in addition to the low-overhead division of labor capabilities, is another advantage of this model.

Assignment mechanisms. Even if users have agreed a priori on a division of sub-tasks within a search session, sometimes one partner might come across something relevant to the other. The ability to "assign" certain sites to other users (*i.e.*, when they log into the search application they see a message with pointers to the sites that other users have suggested they examine) is useful for facilitating on-the-fly division of labor.

3.2 Extensions for Co-located Collaboration

In addition to remote collaboration, a large percentage of our respondents also reported engaging in co-located collaborative searching, such as at a group meeting when a search window is projected onto a large display and several group members suggest keywords, or, more commonly, when one person looks over the shoulder of another and suggests keywords or links to click while their partner controls the mouse and keyboard.

An interesting avenue for further work would be to consider the design requirements and possibilities for supporting a more interactive, engaging, and efficient model for co-located collaborative search scenarios. For instance, mobile devices (such as cell phones, which are nearly ubiquitous among many user populations) that each person is likely to possess could be leveraged to allow increased parallelization of the search process, perhaps by allowing users to search on their mobile devices and collect useful links on a shared display, or allowing the person controlling the "main" computer's input devices to enter a keyword that pushes search results pages out to each of the mobile devices for the other user(s) to inspect in parallel.

4. RELATED WORK

4.1 Studies of Search Practices

This research adds to a body of literature that uses surveys, observation, interviews, and log data to understand the way that people use the web, and, in particular, the way that they conduct web searches.

Several researchers have developed classifications for web search behavior as a result of such studies. For instance, Broder [3], on the basis of user surveys and log analyses, classified web search tasks into three categories: navigational, informational, and transactional. Rose and Levinson [14] did a log analysis study and developed a related taxonomy of navigational, investigational, and resource queries. Their study also found that 35% of queries were not only informational, but exploratory (*e.g.*, asking questions, general research). The prevalence of exploratory search tasks suggests that there is value in considering how these types of searches can be better supported.

White and Drucker [20] found that log analyses could be used to classify users according to two types of search behavior: navigation and exploration. They suggest that different interface designs might benefit these two different searching profiles. We also advocate changing interface design to support explorer-like tasks; our survey data complements White et al.'s log study by offering a more detailed picture of the types of exploratory tasks that could use additional user interface support.

Our survey population consisted of knowledge workers at a large corporation. Sellen et al. [16] conducted an interview study of knowledge workers to learn about how they used the web; however, their study focused on web use in general, not search specifically. Aula et al. [1] also studied general search strategies of experienced web users, and, like us, noted that users frequently employed the Passive Storage/Passive Retrieval strategy of attempting to remember what keywords had previously been used in order to re-conduct a search.

4.2 Re-finding Information

Jones et al. [9] conducted an interview study to understand how people get back to information they've previously found on the web. Some of the questions in their survey relate to our findings on methods people reported using to re-gain the context of a web search after a time gap or device switch. For example, Jones et al.'s study also found that users were not proactive about marking websites for easier re-finding: they reported that 70% of their interviewees used a "do nothing" method for saving information found online.

Tauscher and Greenberg's study [17] found that 58% of web pages people visited were actually re-visits. Systems for helping people re-find previously found information, such as personalized search [18], seek to support revisitation of previously found sites. Our proposed interface designs support "collaborating alone" after a time gap or device switch, such as through providing persistent search session state. Thus, we also add to the body of work on re-finding by proposing means for reducing unnecessary duplication of effort in re-executing previous searches.

Some proposed interfaces reduce the need to re-find information by helping users maintain context during complex search tasks; For instance, SearchPad [2] allows users to explicitly flag a website for inclusion in a workspace to maintain context during complex search tasks. However, our proposals go beyond this model of context to include implicit persistence of sites visited, keywords entered, and a variety of other metadata, in addition to allowing multiple users to participate in and share the process and products of a search.

4.3 Interruptions

Several studies, such as [4] and [8], document the fact that knowledge workers are interrupted and switch tasks with high

frequency. The ubiquity of interruptions, in addition to self-motivated task suspension or switching, is likely a factor in causing so many of our respondents to have search tasks that span gaps in time. A web search interface with a persistent model would likely facilitate the quick regaining of context upon task resumption after interruptions; this is an issue that merits further investigation.

4.4 Cooperation During Search

Our survey highlights the surprisingly high level of cooperative web search behaviors users engage in, despite the fact that these behaviors are not directly supported by search user interfaces. A few prior studies have also found evidence of cooperation during searching activities. For example, Large et al.'s fieldwork [11] found that elementary-school students often collaborate during information-seeking tasks, both due to limited computer resources in classrooms and due to group-based assignment pedagogies. Twidale's study of college students [19] identified several ways in which students collaborated while using the library's database searching terminals. Fidel et al. have done ethnographic work, studying collaborative information-finding behaviors in work settings, although their research focuses on many types of information-finding rather than specifically on web search [5]. Hansen and Järvelin have specifically studied the patent handling process, observing high levels of collaboration on information-retrieval aspects of that task [7].

Researchers have introduced some device- or domain-specific technologies for facilitating collaborative search. For instance, TeamSearch [13] allows up to four co-located users to use a digital tabletop to search through tagged photo collections. Maekawa et al.'s system [12] allows co-located users with Web-enabled mobile phones to improve their visual search efficiency by dividing a web page up among each of their screens. Krishanppa's system [10] allows two remote users to collaborate while searching a medical database by adding a dedicated textual chat facility to the UI, as well as integrating the ability to send information about useful search results to the remote partner. In this paper, we have provided recommendations for the design of a general-purpose web search interface for supporting multiple remote or co-located users in collaboration during both the search and sensemaking processes.

A few commercial products address subsets of our design recommendations; however, none of these products provides an integrated cooperative searching and sensemaking experience, as we have proposed. For example, Google Notebook allows a user to manually designate portions of a website for storage in a persistent collection; multiple users can edit a single "notebook" document. Searching on Google while logged in with a special account allows a user to access a history of query terms typed; however, this list is chronological and cannot be organized according to search topic/session. The Live Messenger IM client provides a "shared search" feature: conducting a web search using the client allows the returned URLs to be a sent to a chat partner. Tagging sites like del.icio.us allow users to store bookmarks online, and to share tags and bookmarks with others.

5. CONCLUSION

We presented the results of a survey of 204 knowledge workers at a large organization. This survey data revealed a variety of web search habits and needs that are not well-supported by current search interfaces, such as the ability to resume a search after elapsed time and/or after switching computers, and the desire to collaborate with others in order to search the web. Based on this data, we formulated design guidelines suggesting key features for supporting these behaviors.

Hopefully, these survey results and design recommendations will inspire the creation of interfaces that support collaboration on web searches. Such systems would open interesting avenues for research within the information retrieval community, such as exploring how work on personalized search might be extended to apply to groups of users, or how query logs and clickthrough data from a group of collaborating users might impact result rankings for a joint search task.

6. ACKNOWLEDGMENTS

We thank Eric Horvitz, Susan Dumais, and Jaime Teevan for insightful conversations about this work. We also thank Ed Cutrell, Jonathan Grudin, and Dan Morris for assistance revising survey questions.

7. REFERENCES

[1] Aula, A., Jhaveri, N., and Käki, M. Information Search and Re-access Strategies of Experienced Web Users. *Proceedings of WWW* 2005, 583-592.

[2] Bharat, K. SearchPad: Explicit Capture of Search Context to Support Web Search. *Proceedings of WWW 2000*, 493-501.

[3] Broder, A. A Taxonomy of Web Search. *ACM SIGIR Forum*, 36(2), 2002, 3-10.

[4] Czerwinski, M., Horvitz, E., and Wilhite, S. A Diary Study of Task Switching and Interruptions. *Proceedings of CHI 2004*, 175-182.

[5] Fidel, R., Bruce, H., Pejtersen, A.M., Dumais, S., Grudin, J., and Poltrock, S. Collaborative Information Retrieval. *The New Review of Information Behaviour Research*, 1(1), 2000, 235-247.

[6] Greenberg, S. and Roseman, M. GroupWeb: A WWW Browser as Real Time Groupware. *Proceedings of CHI 1996 Conference Companion.*

[7] Hansen, P. and Järvelin, K. Collaborative Information Retrieval in an Information-Intensive Domain. *Information Processing and Management*, 41(5), 2005, 1101-1119.

[8] Iqbal, S. and Horvitz, E. Disruption and Recovery of Computing Tasks: Field Study, Analysis, and Directions. *Proceedings of CHI 2007*, in press.

[9] Jones, W., Bruce, H., and Dumais, S. How Do People Get Back to Information on the Web? How Can They Do it Better? *Proceedings of Interact 2003*, 793-796.

[10] Krishnappa, R. Multi-User Search Engine: Supporting Collaborative Information Seeking and Retrieval. *Master's Thesis, University of Missouri-Rolla*, 2005.

[11] Large, A., Beheshti, J., and Rahman,T. Gender Differences in Collaborative Web Searching Behavior: An Elementary School Study. *Information Processing and Management*, Vol. 38, 2002, 427-433.

[12] Maekawa, T., Hara, T., and Nishio, S. A Collaborative Web Browsing System for Multiple Mobile Users. *Proceedings of PERCOM 2006*, 22-35.

[13] Morris, M.R., Paepcke, A., and Winograd, T. TeamSearch: Comparing Techniques for Co-Present Collaborative Search

of Digital Media. *Proceedings of IEEE Tabletop 2006*, 97-104.

[14] Rose, D. and Levinson, D. Understanding User Goals in Web Search. *Proceedings of WWW 2004*, 13-19.

[15] Russell, D., Stefik, M., Pirolli, P., and Card, S. The Cost Structure of Sensemaking. *Proceedings of CHI 1993*, 269-276.

[16] Sellen, A., Murphy, R., and Shaw, K. How Knowledge Workers Use the Web. *Proceedings of CHI 2002*, 227-234.

[17] Tauscher, L. and Greenberg, S. Revisitation Patterns in World Wide Web Navigation. *Proceedings of CHI 1997*, 399-406.

[18] Teevan, J. Supporting Finding and Re-Finding Through Personalization. *Doctoral thesis, Massachusetts Institute of Technology*, 2007.

[19] Twidale, M., Nichols, D., and Paice, C. Browsing is a Collaborative Process. *Information Processing and Management*, 33(6), 1997, 761-783.

[20] White, R. and Drucker, S. All Searches Are Not Conducted Equal: Investigating Behavioral Variability in Web Search. (under review).