

How Knowledge Workers Use the Web

Abigail J. Sellen, Rachel Murphy

Hewlett Packard Research Labs
Filton Rd., Stoke Gifford
Bristol UK BS34 8QZ

Abigail_Sellen@hp.com, Rachel_Murphy@hp.com

Kate L. Shaw

Psychology Group,
Aston University, Aston Triangle
Birmingham UK B4 7ET

katelshaw@hotmail.com

Abstract

We report on a diary study of how and why knowledge workers use the World Wide Web. By examining in detail a complete two-day set of Web activities from each of 24 people, we construct a framework with which to describe the different tasks knowledge workers undertake. By looking at the characteristics of each type of activity, we can see how certain activities are unsuited to particular kinds of technologies (e.g., mobile devices); how Web tools might be incrementally improved; and how we might better support knowledge workers' Web tasks in the future.

Keywords

World Wide Web, knowledge workers, diary study, taxonomy, mobile technology, appliances

INTRODUCTION

The power of the Web and its enabling infrastructure, the Internet, continues to change the way we work and live. As its use spreads inexorably across the globe, one cannot help but feel that its true potential is still yet to be realized. At the same time, indications are that experienced users' interest in the Web for anything more than a small set of mundane purposes is waning [9]. Perhaps this is unsurprising given the significant lack of innovation in this area. What most users think of as "the Web" continues to be very much shaped and limited by the interface through which most of us use it—namely, through Web browsers. These have changed very little over the past eight years.

Unleashing the power of this technology may therefore require changing the way we interact with the Web. Especially in the research and design community, signs are that new Web-based devices, services, tools and enriched infrastructures may be on the horizon. We are beginning to see new kinds of applications and devices connected to the Web (e.g., [17]), new methods of searching and extending Web information (e.g., [1]) and new ways of accessing Web information, for example through physical objects and locations (e.g., [14]). At the same time, technological

advances mean more devices are now internet-enabled, mobile phones being an obvious example.

In our own laboratory, we too have been studying ways in which the power of the Web might be exploited in innovative ways. Central to this endeavor is taking a user-driven approach rather than a technology-driven one. This means we aim to understand what new kinds of technological offerings *users would value* rather than simply asking what the technology could provide.

Our approach is to begin to answer this question by looking at how the Web currently finds its place within people's lives. In other words, this is to ask: What are the different kinds of activities that people use the Web for, and what are their characteristics? What sorts of activities does the Web naturally support well and where does it fall down? What is the context of use of the Web and what other kinds of tools and activities are carried out in conjunction with it? By exploring the answers to these questions, we hope to show that we can begin to look at new uses and new incarnations of Web-based devices and applications. At the very least, we hope to show that such an understanding of the range and nature of what people currently do with the Web provides a reality check for our notions and prejudices of what the Web is used for, and for judging whether the technologies we are currently developing make sense in light of this understanding.

EXISTING RESEARCH

There is a growing body of research on Web usage. Much of this can be found in the Library and Information Science literature (e.g., [2], [13]). However, within this field as well as within Human Computer Interaction, research mainly falls into two camps. On the one hand there are studies that rely primarily on questionnaire, rating scale or interview data. Many of these are large scale studies which ask users to make general statements about how often they do certain kinds of Web activities, how they feel about different aspects of the Web and so on (e.g., [15]). These aggregated statistics are then used to draw conclusions about social, cultural and demographic trends.

On the other hand, and using a different set of techniques, there are studies that rely mainly on data collected on-line. Such "click-stream" studies collected on a large scale can reveal emergent patterns in Web use, showing for example navigational patterns or enabling mathematical modeling of Web use (e.g., [7,10]). Other studies using on-line data are of a task analytic nature, describing the moment-to-moment

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

CHI 2002, April 20-25, 2002, Minneapolis, Minnesota, USA.

Copyright 2002 ACM 1-58113-453-3/02/0004...\$5.00.

behavior of smaller sets of users performing different kinds of Web tasks (e.g., [4,6]).

For the purpose of designing new Web-based technologies, neither approach is sufficient in itself. Interview and questionnaire studies rely on people’s perceptions and memories, ask people to make broad generalizations about what they do and think, and provide no data on actual use. On the other hand, click-stream studies usually have little to say about the goals and motivations of people using the Web and about the larger, day-to-day context in which these activities are carried out. This is particularly the case with aggregated click-stream data, but it is also a problem with the task analytic studies.

There are two exceptions to this. One is a taxonomic analysis by Morrison et al. [11] looking at how and why people do different Web activities. However, this was based on data collected through a Web survey asking people to submit examples of Web use that helped them in taking action or making decisions. As such, this analysis, while shedding some light on people’s goals and strategies, biases the data toward particular kinds of events. In addition, because the method relies solely on self-report, it does not capture as much contextual detail as one might like. Another study by Wei Choo et al. [16] uses a combination of click-stream and questionnaire data to analyze the activities of 34 knowledge workers, but again, only a selection of tasks are fully analyzed and the context in which they were carried out is not examined in detail.

Our Approach

The purpose of this study, then, was to provide some data on a comprehensive, detailed sample of Web activities. To do this, we chose a combined diary and interview methodology. This allows us to look at the actual activities occurring in a “slice” of people’s everyday lives, and to unpack the detail of those activities through in-depth interviews. While such a methodology cannot claim to generate as representative a sample of data as larger scale studies, it can provide important input for inspiring and shaping new design ideas.

With regard to the scope of the study, we chose to study knowledge workers. Knowledge workers are paid to transform knowledge. In doing so, they tend not to work in a routine way but often employ ad hoc methods for the project at hand. We reasoned that this group of people might reveal an interesting diversity of ways in which the Web would play into their work. Accordingly, we also decided to look at activities during working days but not to exclude non-work activities in the workday, this being an interesting aspect in itself. Finally, we decided to exclude email activities in our data since whether participants chose to read email through the Web had little to do with the functionality of the Web but rather had to do largely with organizational and historical factors.

METHOD

Participants

Twenty-four knowledge workers were recruited outside of HP Labs. A knowledge worker was defined as someone “whose paid work involves significant time: gathering, finding, analyzing, creating, producing or archiving information.” Our definition of “information” was also quite broad meaning anything from documents, policies, plans, and presentations to drawings, designs and graphics. We also recruited people who accessed the Web at least four times in a typical working day (excluding email).

The result was a diverse group of individuals (Table 1) ranging in age from 22 to 55 and with between 2.5 and 10 years experience using the Web. Most were office-based although 4 people worked primarily from home and one was mainly mobile (away from her desk 75% of the time).

Table 1. Summary description of participants.

No.	Job Title	Age	Yrs on Web	Main Workplace
1	Lead software developer	34	7	Office
2	Information scientist	34	8	Office
3	Magazine production editor	29	3	Office
4	Financial advisor	33	4	Office
5	University lecturer	44	10	Office
6	Trade marketing manager	29	2.5	Office
7	Optical design engineer	30	6.5	Office
8	Magazine editor	36	7	Office
9	Usability engineer	35	10	Office
10	Senior lecturer	43	4	Office
11	Housing asset manager	42	4.5	Office
12	Government contracts manager	32	5	Office
13	Building historian	55	3	Home
14	Government IT advisor	41	7	Office
15	Production assistant for the BBC	34	6	Office
16	Assistant producer for the BBC	35	3	Office
17	Magazine editor/journalist	52	4	Home
18	Senior lecturer	49	6	Office
19	Technical advisor for images	22	7	Office
20	Broadcast journalist	26	7	Office
21	Marketing consultant	45	3.5	Home
22	Marketing consultant	39	4	Home
23	Child psychologist	50	6	Mobile
24	Free lance journalist	45	7	Office

Procedure

All 24 participants were informed that we would ask about the Web activities they undertook in the course of two consecutive working days. At the end of each of the two days, they were interviewed in front of their PCs with their normal Web browser opened to reveal their history list. Participants were first asked to tell a “story” about each activity they carried out, meaning they were asked to describe the activity in terms of why they carried it out, what happened during the activity and any events leading up to or following on from the activity. The definition of an activity was left up to them, and often involved several of the websites in their history list. A separate form was filled

out for each activity to record these free-form descriptions. In addition, for each activity, a series of supplemental questions was then asked:

- How did you initiate this activity?
- Was it a work or non-work activity?
- Was there a goal? (If yes, did you achieve it?)
- Was there a specific question (and if so what was it?)
- Was the activity initiated, executed and completed in one sitting, or was it spread out over time?
- Was this a routine activity or unique event?
- Was there any collaborative aspect to the activity?
- Was anything printed out?
- What were the good or bad aspects of the way the Web supported this activity?

Three numbers were also obtained: Participants were asked to rate the success of the activity in achieving their goal (if any); to rate the overall important or significance of the activity; and to estimate the amount of time spent on each activity (using on-line data when available). Ratings were given on a 10-point scale; time estimations were given to the nearest minute. These three numbers formed the basis of many of our quantitative analyses. Interviews typically took about an hour.

RESULTS AND ANALYSIS

As might be expected, both frequency and duration of Web use varied considerably across the sample of knowledge workers. In all, 295 activities were recorded for the 24 participants – a mean of 12.3 activities per person with a standard deviation of 6.1 across participants. On average, then, they carried out about 6 Web activities a day, but this ranged from 1 to 15 activities in a day. In terms of time, participants spent a mean of 91 minutes a day on the Web but again this varied considerably. For example, one person spent only 5 minutes on the Web on day two, while another spent 5 hours streaming radio on each of both days.

Activity Classification

More interesting than these overall statistics are the details of the kinds of activities they carried out. On close inspection of the data, clusters emerged, each having common characteristics, goals and purposes. We agreed upon and sorted the activities into six such categories:

- **Finding:** Using the Web to find something specific. In this case, searching is goal-oriented and very well defined: E.g., Finding a fact such as a phone number, spelling or product name; a set of facts such as a list of ingredients for a recipe, or list of train times; or a virtual product or products such as a document, software, map, or image.
- **Information Gathering:** Less specific than “Finding”, but using the Web to purposefully research a specific topic for various reasons. E.g., Gathering information in order to compare, choose or decide about something (such as buying products or looking for jobs); in order

to supplement a future task (such as collecting background information to write a document, or to prepare for a meeting); or in order to be inspired or get ideas.

- **Browsing:** Going to sites out of personal or work-related interest with no specific goal in mind but rather to be informed, stay up to date or be entertained. E.g., Browsing through a newspaper or magazine, following an interesting link, or checking to see what’s new on a hobby-related site.
- **Transacting:** Using the Web to execute a transaction securing future products or services: E.g., Making a bank transfer, paying a bill, ordering a physical product, or filling out questionnaires.
- **Communicating:** Using the Web in order to participate in chatrooms or discussion groups.
- **Housekeeping:** Using the Web to check or maintain the accuracy and functionality of Web resources. E.g., Checking that information on a Web site is up to date, that links are working properly and so on.

For the most part, each activity clearly fell into one of the six categories. However, for 13, the activity consisted of elements of two categories. (For example, Finding was on three occasions followed by buying the item found, which put these also in the Transacting category.) For compound activities, task duration was divided evenly and allocated to both categories for time-based analyses. For analyses based on frequency, each compound event was counted as two events (one in each category).

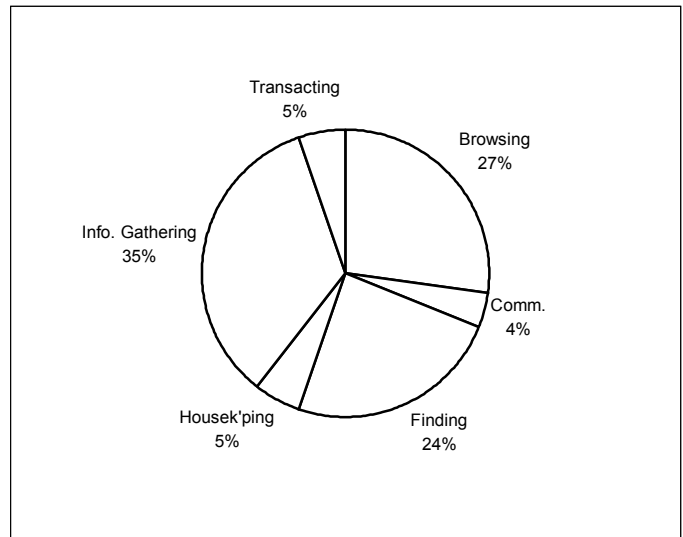


Figure 1. Frequency of different kinds of Web activities expressed as a percentage.

Overall Description of the Data

Figure 1 shows the frequency of these different activities and Figure 2 shows their average duration. From these data we can see that Information Gathering activities were the most frequent and also of second longest duration, taking on average about 23 minutes. Both Finding and Browsing were

also quite frequent but tended to be of a much shorter average duration (just under eight minutes). The last three activity categories, Transacting, Communicating and Housekeeping were relatively infrequent. Note also that Housekeeping activities, although not common, tended to be time consuming, taking over 40 minutes on average.

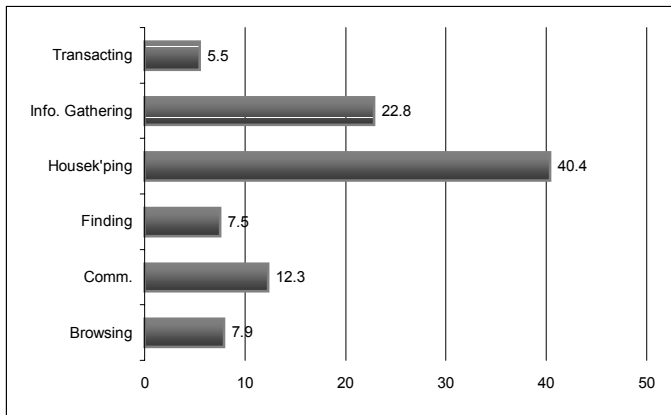


Figure 2. Mean duration (minutes) of different kinds of Web activities.

Both frequency and length of these different activities tell us what kinds of tasks the Web was used for most, as well as which ones these knowledge workers spent the most time on. But what kinds of activities did they feel were the most important? The results show that Browsing activities were rated as less important than the more goal-oriented categories (Table 2). This was also reflected in participants' comments. Many Browsing activities were described as things done when taking a break from a task, or as something done between tasks. This can be contrasted with Finding and Information Gathering activities that were often described as central to some larger task with a specific goal or set of goals.

Table 2. Mean importance ratings of activities.

Category	Mean Imp.*	Results of Pairwise Tests	F statistic
Browsing	5.4	see below	see below
Communicating	5.8	no sig. diffs	--
Finding	6.9	>Browsing	F(1,152)=10.5, p<.001
Housekeeping	7.6	>Browsing	F(1,96) = 7.0, p<.01
Info. Gathering	6.8	>Browsing	F(1,183)=14.5, p<.001
Transacting	7.3	>Browsing	(F (1,96) = 5.0, p< .028

*1 = "very unimportant; 10 = "very important"

A final point to make is about the extent to which these Web activities were collaborative. As we will discuss, asynchronous collaboration played a role in many of the activities (mainly through email). Remote collaboration too took place via telephone from time to time. However use of the Web itself was mainly a solitary activity although, on some occasions, Web pages were shared or jointly viewed with others who were co-present.

Activity Characteristics

A closer examination of these categories shows that each has associated with it a distinguishing set of characteristics. For the sake of brevity (and because they are the most interesting) we limit further analysis to the categories of Finding, Information Gathering, Browsing and Transacting:

Finding

This category involved activities where goal-oriented and focused questions were asked: E.g., "What is the reference for this publication?" "How do I get from Bath to Slough tomorrow?" "What is the curvature of this optic lens?" "How do I go about registering for this conference?"

Many different kinds of information were sought. Most were *facts* such as names, dates, phone numbers, quotations, weather forecasts, procedures, prices, product names, financial data, schedules, and directions. A few were *images* such as maps, photographs and Webcam pictures. A few were instances of finding and downloading *virtual products* such as software, screensavers, images and documents. Most of the information sought, however, ultimately took the form of single words, phrases, or lists of text.

The participants were quite experienced at this sort of searching, and, overall, were quite successful (achieving or partially achieving their goal in 78% of cases). Perhaps too because of their experience, most of these sorts of activities were both short and self-contained. In other words, for 90% of cases, they were initiated, executed and completed without interruption.

Having said that, finding answers was not necessarily a simple activity. This kind of searching often involved multiple keywords, following many links, and scanning through many different documents and sites. In a few cases, participants resorted to the telephone to "speak to a real person" when queries were not answered adequately, or when information was too hard to find.

The telephone was often useful, then, as a quicker way of obtaining accurate information customized to the particular question. But it was also useful because there was sometimes uncertainty as to the trustworthiness of the information provided. Key here was the degree to which users perceived sites to be up to date since many of the facts that were sought (such as train schedules or prices) were often changed. One participant said he didn't trust "brochure sites" or sites perceived to be static and out of date. Another said he didn't trust sites where it appeared he was not talking direct to the organization but rather that the site had outsourced their Web services.

The issue of trust was also critical for those whose jobs very much depended on the accuracy of their knowledge. For example, the journalists, producers, marketing consultants and historian raised this as an issue in assessing the quality of information returned by the Web. Discussion often centered on the importance of using branded sites such as the Encyclopedia Britannica, or sites associated with official societies or universities. Such resources were usually

bookmarked and used more frequently than others. This is reflected in the fact that, overall, general search engines were used only 19% of the time. Much more common was the use of bookmarked or familiar sites specialized around particular kinds of knowledge. This finding has important implications for design, which we shall come back to.

A final point to note is that this kind of information seeking can also be viewed in the context of the larger tasks they were part of. Most searches (86%) led to further activities. Listed in order of decreasing frequency, once information was found, participants wanted to:

- *Refer back to it* ($n=19$). Information such as maps, procedures, and lists needed to be kept, usually only temporarily, so that they could be referred back to. Often this was achieved through printing.
- *Incorporate it into a document* ($n=13$). This included many types of information such as addresses, spellings, references, facts, and images.
- *Use it to make a phone call* ($n=8$).
- *Monitor it* ($n=6$). Sites or specific information needed to be monitored for change such as price changes, and updates to sites.
- *Email it* ($n=6$). When information needed to be shared.
- *Download it* ($n=4$).
- *Buy it* ($n=3$).

In terms of design implications, it is useful to consider these follow-on activities, as we will come to later.

Information Gathering

Information Gathering, or collecting information around a particular topic, differed from Finding in that many of these activities involved a *set* of questions to be answered (e.g., “Who are the key players in this market? What other products do they offer? Where are these people located?”), or in that the questions were quite open ended (e.g., “What job opportunities are out there?”). In other instances, it was a case of not knowing in advance how much information there was to be gathered (e.g., “Has this particular story been covered yet?”). Such activities often involved comparing and contrasting information across sites or organizations (e.g., “How does my insurance policy compare with others?”). In many cases, however, although these knowledge workers could sometimes specify a high level question they were posing, often they could not. Rather, they said they wanted general information about a topic (a country, a product, a company, a person and so on).

Such activities were mainly project-driven. Interestingly, as with Finding, such tasks were usually initiated through bookmarked specialist sites. By contrast, general search engines were used only 29% of the time.

Here again, the use of trusted sites was a key issue. As the information scientist put it, one needs to know about the *quality* of information—certain sites provide better quality information than others. Part of this knowledge worker’s job was to know and assess which sites provided the best

quality information and which did not. Thus, an important phase of his work was when he would engage in “scoping searches”: time dedicated to evaluating sources, saving the valuable ones to folders. Later, he would revisit these to focus in on the documents buried within them. Other knowledge workers too, with quite different jobs, talked of the need to gather information from trusted, reputable sources. Such resources were bookmarked and used more frequently than others. When information came from an unknown source, it would be checked against other sources such as reference books.

While sometimes the accuracy of information was a significant issue in using the Web to gather information, this was not always the case. Some information gathering activities were for the purpose of inspiring ideas (e.g., journalists looking to see what the hot stories were and following up leads.) Others were simply to provide background information on a topic area. As one of the marketing consultants put it, the Web could be very useful in “building a picture” of an organization, topic or person. The telephone was often used after these background searches, to carry out interviews or set up meetings.

Whatever the purpose, information gathering activities could be quite time-consuming and complex. More often than not, such meta-level tasks encompassed navigating multiple links and sites. It was clear that this process was very reliant not only on complex navigational patterns, but also on scanning and skimreading large amounts of material to assess its relevance. Another characteristic of these activities was that they often unfolded over time, sometimes spread out over days or even weeks. We found that 40% of these activities were not completed in a single session, being interrupted either due to other demands or because of the amount of time required to complete them.

Because of this, participants used various ways of saving the interim results of their Web activities. Some used virtual folders (either on the desktop or bookmark folders) to collect together links associated with different topics or projects. Others printed out their findings, including search results, tables of contents and home pages. When asked why they chose to print rather than use electronic folders, some participants complained of the ever-increasing numbers of electronic folders they had to search through and manage. However, a more fundamental concern was that interim search results would not be preserved or that information would change in the future, making it hard to get back to where they had been. By printing, results could be quickly to hand when needed, making it easier to resume a task.

Browsing

Browsing differed from the first two categories in that it was not generally goal-driven. Rather, knowledge workers used the Web to “see what’s new”, “keep up to date”, or even “waste time”. Some participants dismissed these activities as unimportant or purely for the sake of entertainment. Others talked of the importance of keeping up with events in a work-related field, in a home country, or

in hobby-related sites such as science fiction, gardening and football sites.

Over half of these activities (57%) were routine in that they were carried out either daily or weekly. In these cases, favorite sites either were bookmarked, entered from memory, or reached through automated email lists. The remaining 43% of these activities were one-off browsing events, either as the result of emailed links from friends and colleagues, or through serendipitously discovering and following interesting links while at other websites.

Such activities tended to be both short in duration and also self-contained (on only four occasions were they spread over time or interrupted). Information was also rarely printed out or saved in a virtual folder, and there were usually no follow-on activities except occasionally to bookmark or email an interesting item to someone else.

The kind of information browsed was here, as in all other categories, primarily visual, usually involving scanning a pageful of text and graphics. However, some of these items were of a smaller format such as photos, the weather, sports scores, jokes, and stock prices. Participants also heavily relied on headlines and summaries.

In addition, unlike the other categories, about 10% of events involved audio, and a couple of events involved video. Indeed, two activities involved using the Web to stream music; in another, a different participant listened to the radio through a website. In some sense such activities may be thought of as similar to visual browsing in that they are not necessarily goal-driven and are for the purpose of interest or entertainment. However, being auditory in nature, they can be carried out as a background activity.

Transacting

The last category we will describe is that small subset of activities in which these knowledge workers used the Web for executing transactions. Though clearly goal-driven, such activities did not necessarily involve answering questions or seeking information: most involved quite directed, straightforward attempts to secure a service, order a product, or manage money. Having said that, questions sometimes accompanied such activities (e.g., "What is my bank balance?"). Searching behavior too sometimes was a precursor to executing a transaction (e.g., looking for specific products to buy).

About half of these activities were transactions that these experienced Web users had undertaken before. Perhaps for this reason, the issue of security arose only once as a concern. Most negative comments were about sites taking too long to respond or requesting too much information. Any kind of interruption or delay was not tolerated.

Another issue was that of having to remember too many user identification numbers and passwords. This also arose for other kinds of activities, many kinds of Websites requiring users to register in order to reap the benefits of personalized services. Some participants had systematic ways of dealing with all of their different user names and

passwords (such as using index cards) while others managed them in much more haphazard ways. Whatever the case, the overhead involved was something that participants complained about and avoided when possible.

DISCUSSION AND IMPLICATIONS FOR DESIGN

Perhaps the most obvious point to take from these findings is the extent to which terms like "surfing" or "browsing" the Web gloss over and even misrepresent what people use the Web for. By capturing a sample of real life events and examining them in detail, we can see that there are many different reasons why knowledge workers use the Web and many different behaviors they engage in: some goal-oriented and some not; some centered around focused questions and some not; some short and self-contained and some spread over time. Not surprisingly, these activities involve different patterns of use and types of information. In doing so, they may draw on other resources (documents, technologies, and people). All of these factors contribute to a better understanding of what knowledge workers do with the Web, how they do it and why they do it.

At the very least, this should cause us to question assumptions about Web use that fail to specify what *kind* of use we are talking about, because, as we have seen, such behaviors have their own distinct characteristics. So, if a mobile phone operator says they are going to support Web use, the important question is then "What *kind* of Web use?". If they are not thinking about particular kinds of activities, chances are they are not thinking about how best to support them. This research provides a way of talking about Web-based activities and a framework for understanding what the differences between various activities might be.

Perhaps more importantly, can an understanding of current Web use suggest how Web technologies might move forward into the future? On the one hand, we can postulate that knowledge workers use the Web in the way we have described because of its limitations: that is to say, because they are constrained by current browsers, tied to desktop machines, and limited to its current content, architecture, and services. On the other hand, we could surmise that knowledge workers use the Web in this way because these are the activities that they most need to carry out as part of their job, and that they adapt and select from the resources at hand to do so. Whatever stance one chooses to take, we would argue that a deeper understanding of knowledge workers' activities lets us make more educated guesses about what new kinds of technology they might value and what kinds they would not. Let us consider just a selection of what the results suggest:

"Finding" Technologies

With the proliferation of small, portable, wireless technologies, one of the key questions to ask is: What Web activities might be appropriate to carry out on such devices? Clearly, many of the activities falling into this category could be easily displayed on a small screen. Much of the information sought was of a small size and format (words,

phrases, lists, images). In addition, many of the kinds of information sought (e.g., phone numbers, directions, movie times) are those that people on the move might well value.

What the results also tell us is that while users do need access to search engines to carry out Finding activities, they are more dependent on using familiar sites. Therefore, the interface for small devices needs to enable quick access to these trusted resources. What this says in turn is that bookmarking facilities need to be designed mainly for speed of access, not information management. Current browsers with their hierarchical folder systems are designed to do both. Redesigning for speed, not information management, might result in a radically different set of features. For example, this would mean minimizing the number of actions necessary to get to a site (e.g., getting rid of menus), and might point to different methods of input such as speech input.

Another important implication is that there is a set of follow-on activities that users might want easy access to after information is found. Especially for small mobile devices, one can envision a small, limited set of functions that would be useful to apply once the desired information is located. Some examples might be:

Temporary save and display. The need to frequently refer back to information such as directions, maps, lists, and procedures during a journey or a task implies that a single button facility to display a temporary store of information would be valuable. Users could save found information and instantly access it without having to connect to the Web again, and without cumbersome navigation through menus.

Easy calling. The fact that phone numbers were often sought from Web sites and calls made in conjunction with Web activity points to the value of: instantly highlighting phone numbers on Web sites; allowing one-button calling from sites from a Web-enabled phone; and supporting quick storage of phone numbers from Web pages.

Automatic monitoring. Another function that might be useful especially on portable or wearable Web-enabled devices, would be the ability to “watch” or monitor any Web site or selected portion of a Web page for change. There are currently services which will do this for stock prices, but this could be a much more general kind of feature, for example allowing users to monitor airline prices, changes in Web-cam images, or indeed any update to a site.

As a final note of caution, all of the above suggestions depend on extrapolating from fact finding at the desk to fact finding on the move. One way of ensuring this makes sense is to study and understand the mobile worker, something we have been undertaking in other work [12].

“Information Gathering” Technologies

Unlike Finding activities, the results show that Information Gathering activities would be, for the most part, entirely unsuitable for supporting on small screen devices in mobile settings. Unlike Finding, these sorts of research activities

often involved scanning of large sections of text, complex navigation, information management, and archiving. Thus it is unlikely that any small screen device could successfully support the whole range of what knowledge workers would need to do.

The results also suggest that such activities may be hampered even on large screen devices in settings away from the desk (i.e., laptops) because of the need to rely on other resources and infrastructure to support meta-task activity. In other words, there was a need to check information against other sources (such as books, paper documents, and the knowledge of colleagues) and to use paper folders, printouts, and notebooks to preserve and collate information in the face of tasks that were often protracted and interrupted. On-line Web tools might be improved to help address these problems, however. E.g.,:

- *Better tagging of information.* Tools to help knowledge workers assess trustworthy information would alleviate some of the need to cross-check information. For example, peer reviewed sites, or sites verified in other ways, might be consistently tagged as such. There are also guidelines site designers can follow to increase the credibility of their sites [8]. In this study we saw that perceptions of when a site was last updated and whether it appeared to be connected directly to a physical organisation were two important issues.
- *Better search tools.* Search facilities could be improved in finding trustworthy information by, for example, giving priority to sites which help pull together, assess or “scope” the validity of other sites. Such meta-search facilities might also perform services such as tracking the use of sites by people with similar professional profiles. Current search engines of course employ their own kind of weighting criteria but these remain opaque to the user. In future, search engines could differentiate themselves by making these criteria visible.
- *Webscrapbooks.* The need to save information not just in terms of URLs, but also in terms of a much wider range of information (selected pieces of text, search results, graphics, etc.), implies that more flexible ways of managing information would be valuable. Rather than saving to hierarchical folders, users could construct mixed-media scrapbooks into which they clip and store information from Web pages for particular projects. At the same time, these scrapbooks could be dynamic. For example, intelligent Web services or agents could make further recommendations around its theme. They could also monitor some of the key sites visited and notify the user of updates or changes. This is, in effect, a more specialized version of Xerox PARC’s notion of Webbooks [5].
- *Better history functions.* Such scrapbooks could also help users keep information in a kind of holding pattern when Web tasks are left incomplete or are interrupted. A “history” function within Web scrapbooks could

allow users to see the state of their work at various stages, allowing them to retrace their steps and also to return instantly to the state in which work was left.

CONCLUSION

We end by considering what the results have to say about the notion of a general browser. We have seen that different kinds of Web activities point toward certain kinds of devices, services, applications and infrastructure that best serve those activities. These can be at odds with each other: for example some activities are better supported on small-screen mobile devices while others are more suited to large screen devices in fixed environments rich with supporting infrastructure.

This sort of conclusion is consistent with more extreme arguments for the specialization of Web-based devices, applications and services according to the task at hand (e.g., [3]). This argument holds that users should select the tools best suited to maximize their interaction with information. This might include specifying and streaming different kinds of Web content to different kinds of appliances. For example, as a user I might choose that:

- on-line magazines and newspapers go direct to my printer or e-book;
- shopping sites go to my portable home tablet;
- football scores, stock prices and maps go to my handheld or wearable computer;
- photos and Web-cam pictures go to a wall-mounted display in my kitchen;
- MP3 files go to my internet-enabled audio player;
- video goes to my television or seat-back in the car.

It is easy to see that other advantages could accrue from this approach. Because devices are specialized and pre-configured for particular kinds of information, this avoids long download times for high bandwidth content and the time to acquire the tools to deal with content.

Claiming support for this view on the basis of this study would be to go well beyond the data to hand. Nonetheless, we believe the results do make a strong case for reflecting on what users want to do with the Web, and for more generally taking into account people's motivations, goals and strategies to design better Web software, services and infrastructure to suit these needs. When this happens, it may well be the case that the Web itself recedes into the background, becoming the enabling infrastructure for this vision of the future, not the "thing" one interacts with.

ACKNOWLEDGMENTS

We are extremely grateful for the forbearance of all the knowledge workers who took part in this study. Thanks also to Erik Geelhoed and Sarah Beech for their cheerful help with statistical analysis, and to Richard Harper and Dave Reynolds for comments on the final draft.

REFERENCES

1. Berners-Lee, T., Hendler, J. & Lassila, O. (2001). The semantic Web. *Scientific American*, Vol. 5(1).
2. Bruce, H. (1998). User satisfaction with information seeking on the Internet. *Journal of the American Society for Information Science*, 49 (6).
3. Buxton, W. (2001). Less is more (more or less): Some thoughts on the design of computers and the future. In P. Denning (Ed.), *The Invisible Future: The seamless integration of technology in everyday life*. New York: McGraw Hill.
4. Byrne, M.D., John, B.E., Wehrle, N.S., & Crow, D.C. (1999). The tangled web we wove: A taxonomy of WWW use. *Proc. CHI 99*, 544-551.
5. Card, S, Robertson, G. & York, W. (1996). The Webbook and Web forager: An information workspace for the World Wide Web. *Proc. CHI '96*, New York: ACM Press.
6. Card, S., Pirolli, P., Van Der Wage, M., Morrison, J., Reeder, R.W., Schraedley, P.K., & Boshart, J. (2001). Information scent as a driver of web behaviour graphs: Results of a protocol analysis method for Web usability. *Proc. CHI 2001*, 498-505.
7. Chi, E.H., Pirolli, P., Chen, K., and Pitkow, J. (2001). Using information scent to model user information needs and actions on the Web. *Proc. CHI 2001*, 490-497.
8. Fogg, B.,J, Marshall, J., Osipovich, A., Varma, C., Laraki, O., Fang, N., Paul, J., Rangnekar, A., Shon, J., Swani, P. and Treinen, M. (2000). Elements that affect Web credibility. *Proc. CHI 2000*, Extended Abstracts, 287-288.
9. Harmon, A. (2001). Exploration of the World Wide Web Tilts from Eclectic to Mundane. *The New York Times*, Aug. 26.
10. Huberman, B. (2001). *The Laws of the Web*. Cambridge, MA: MIT Press.
11. Morrison, J., Pirolli, P., and Card, S. (2001). A taxonomic analysis of what World Wide Web activities significantly impact people's decisions and actions. *Proc. CHI 2001*, Extended Abstracts, 163-164.
12. O'Hara, K., Perry, M., Sellen, A. and Brown, B. (2001). Managing information on the move. In Brown, B, Green, N and Harper, R (Eds.) *Wireless world: Social and interactional aspects of the mobile age*. The Hague, Netherlands: Springer Verlag.
13. Pettit, S. (1995). Internet use by UK academic law librarians. *The Law Librarian*, 26(1).
14. Spasojevic, M. & Kindberg, T. (2001). *A study of an augmented museum experience*. HP Labs Technical Report No. HPL-2001-178.
15. UCLA Internet Report (2000). *The UCLA Internet Report: Surveying the Digital Future*. Available at <http://www.ccp.ucla.edu/>.
16. Wei Choo C., Detlor, B. & Turnbull, D. (2000). Information seeking on the web: An integrated model of searching and browsing. *First Monday*, Vol. 5(2).
17. Youll, J. (2001). Wherehoo and periscope: A time & place server and tangible browser for the real world. *Proc. CHI 2001*, Extended Abstracts, 109-110.