Stuff I've Seen: A System for Personal Information Retrieval and Re-Use

Susan Dumais, Edward Cutrell, JJ Cadiz, Gavin Jancke, Raman Sarin, Daniel C. Robbins
Microsoft Research
One Microsoft Way
Redmond, WA 98052 USA
[sdumais; cutrell; jjcadiz; gavin; ramans; dcr]@microsoft.com

ABSTRACT
Most information retrieval technologies are designed to facilitate information discovery. However, much knowledge work involves finding and re-using previously seen information. We describe the design and evaluation of a system, called *Stuff I've Seen (SIS)*, that facilitates information re-use. This is accomplished in two ways. First, the system provides a unified index of information that a person has seen, whether it was seen as email, web page, document, appointment, etc. Second, because the information has been seen before, rich contextual cues can be used in the search interface. The system has been used internally by more than 230 employees. We report on both qualitative and quantitative aspects of system use. Initial findings show that time and people are important retrieval cues. Users find information more easily using SIS, and use other search tools less frequently after installation.

General Terms
Algorithms, experimentation, human factors.

Keywords
Personal information management, user interfaces, user studies, interactive information retrieval.

INTRODUCTION
Most information retrieval tools, like popular web and intranet search engines, are designed to facilitate information discovery. Given a short query, they do a remarkable job of finding relevant materials using a variety of content, anchor text, link and popularity cues. However, much knowledge work involves integrating and re-using information that has previously been created or accessed. For example, writing a presentation or paper may involve some web searching, but it also involves pulling together information from existing information sources like documents, spreadsheets, data analyses, email messages, etc. Studies have shown that 58-81% of web pages accessed were re-visits to pages previously seen [9,23,29].

SIGIR'03, July 28 – August 1, 2003, Toronto, Canada.

Similar re-access patterns have been observed in usage of Unix commands [15], library book borrowing [7], and human memory [3].

We developed a system called ***Stuff I've Seen (SIS***) that makes it easy for people to find information they have seen before. Two key aspects of the SIS design support this. First, the system provides a unified index of information that a person has seen on their computer, whether the information was an email, web page, document, media file, calendar appointment, etc. Today, people have to manage several different organizations of information – e.g., the file system hierarchy for files, the email folder hierarchy for email, favorites or history for web pages. With SIS, all of these sources are integrated into a single index regardless of what form the information originated. Second, because a person has seen the information before, rich contextual cues such as time, author, thumbnails and previews can be used to search for and present information. In contrast, web search results lack personal context, so rank is about the only reasonable alternative for ordering results.

RELATED WORK
Vannevar Bush's vision of memex [8], "*a device in which an individual stores all his books, records, and communications, and which is mechanized so that it may be consulted with exceeding speed and flexibility*" captures the essence of SIS. The details of implementation are obviously very different, but the functionality is much the same as what we have developed.

Jones and Thomas [18] surveyed people's use of new personal information management technologies and found low adoption rates. They speculated that the limited applicability of new technologies, each focusing on a limited conception of personal information, was the main reason for slow adoption. Erickson [12] developed a new personal notebook application. He described field observations of usage patterns as well as the co-evolution of the system and work practices. Our focus was not to create a new information management application but rather to develop a unified search interface to existing content sources.

Several groups have looked at methods for improving access to subsets of information. Malone [21] observed

how people organize paper materials and suggested how computer systems could better support these activities. Barreau and Nardi [4] examined how people manage files on their computer. Whittaker and Sidner [30] described the email overload problem and linked it to difficulties that people have with deferred processing and classification of items. Jones et al. [17] conducted detailed observations of the methods that people use to organize web pages for re-use and developed a functional analysis to show how the techniques people use depend on anticipated re-access needs. Several groups developed systems to improve re-access to web pages, including the use of rich graphical representations [10], integration of back, history and favorites [19], predictive models of information needs [24], and full text indexing of pages [22]. These investigations all focused on improving access to information within a single content type.

Some research systems provide access to more than a single type of personal information. Haystack [1,16] is a personal store that supports annotations and collections. The initial version worked with input from the browser and editor, although extensions to email and files were planned. MyLifeBits [14] is similar to SIS but focuses on multimedia files and support for rich annotations. Lifestreams was developed as a replacement for the desktop metaphor [13]. It provides a single time-ordered stream of electronic information, and supports searching, filtering and summarization.

Several commercial products have functionality to index some personal information. Microsoft's Indexing Services and Apple's Sherlock index files, but do not work with mail stores or the web. Other desktop search applications like Enfish Personal, PC Data Finder, 80-20 Retriever, and Scopeware index both files and email. Some also index web pages, and others extend indexing to enterprise content (that the user may or may not have seen). However, we are not aware of published research on how these systems are used, or how they affect people's work patterns.

We believe that SIS covers a wider range of information sources and file types than the systems mentioned above. More importantly, our focus is on studying user's experiences with SIS, and exploring novel retrieval algorithms and interfaces that capitalize on the user's familiarity with their own content.

STUFF I'VE SEEN (SIS)

Today it is often easier to find information on the web than on your own desktop, email store, or intranet. This is due to both the multiplicity of independent applications used to manage information each with its own organizational hierarchy (e.g., email, files, web, calendar), and to the limited search capabilities in many of them. SIS remedies this problem by providing a unified index across these different information sources. If a user wants to restrict search to a particular source they can, but this is not a prerequisite for finding information.

In addition to the core indexing capabilities, we explored new ranking and presentation ideas in SIS. Because the information is personal and has been seen before, we believe that rich contextual cues such as time, author, thumbnails and previews can be especially useful. Moreover, the local index allows for very fast searching and query refinement.

We report on the design and functionality of SIS, and our experiences in deploying it to hundreds of diverse users.

System Architecture

The SIS application is built on top of a modular MS Search indexing architecture. There are five main components. The *Gatherer* specifies the interface to different content sources in their native format. Files, http, and MAPI are examples of gatherers that are supported in the current prototype. The *Filter* decodes individual file formats (e.g., doc, pdf, ps, html) and emits a character stream for further processing. The *Tokenizer* breaks the stream of characters into 'words' and can also handle additional linguistic processing such as date normalization, stemming, etc. The Gatherer, Filter and Tokenizer components are extensible to handle new data sources, file types and languages. The *Indexer* builds a standard inverted index structure with position information to support quick retrieval. The *Retriever* is the query language for accessing stored information. It supports Boolean as well as best match retrieval on the full text and metadata properties. The best match algorithm is based on Okapi's probabilistic ranking algorithm. The Retriever also allows phrase, wildcard and proximity searches.

All of the SIS components run on the client machine. By default, the users' mail profile, web cache, and personal files including media files are indexed. Other data sources can be added. The index is automatically updated as new mail is received, web pages viewed, or content created.

User Interface

The SIS interface allows users to specify queries and to view and manipulate results. Because SIS works from a local index, query results can be returned very quickly, allowing a highly interactive and iterative query strategy. Contrary to many search interfaces where users specify several properties and then press a button to launch a query, SIS launches its queries whenever any of the filtering widgets in the UI are manipulated or when the user presses return. This allows a user to start broadly and then quickly refine their query by interactively filtering and sorting the results. These interface ideas are related to Belkin et al.'s [5] work on iterative query refinement. They are also similar to Ahlberg et al.'s [2] work on dynamic queries except that SIS works on a large personal text collection with discrete attribute values.

Figure 1 shows the first interface we developed, called the Top View. It is a list view with filters for refining attributes in each column. The query text box is in the upper left hand corner. By default, query words are combined using an AND operator. Users can specify other Boolean operators, a fuzzy matching alternative in which

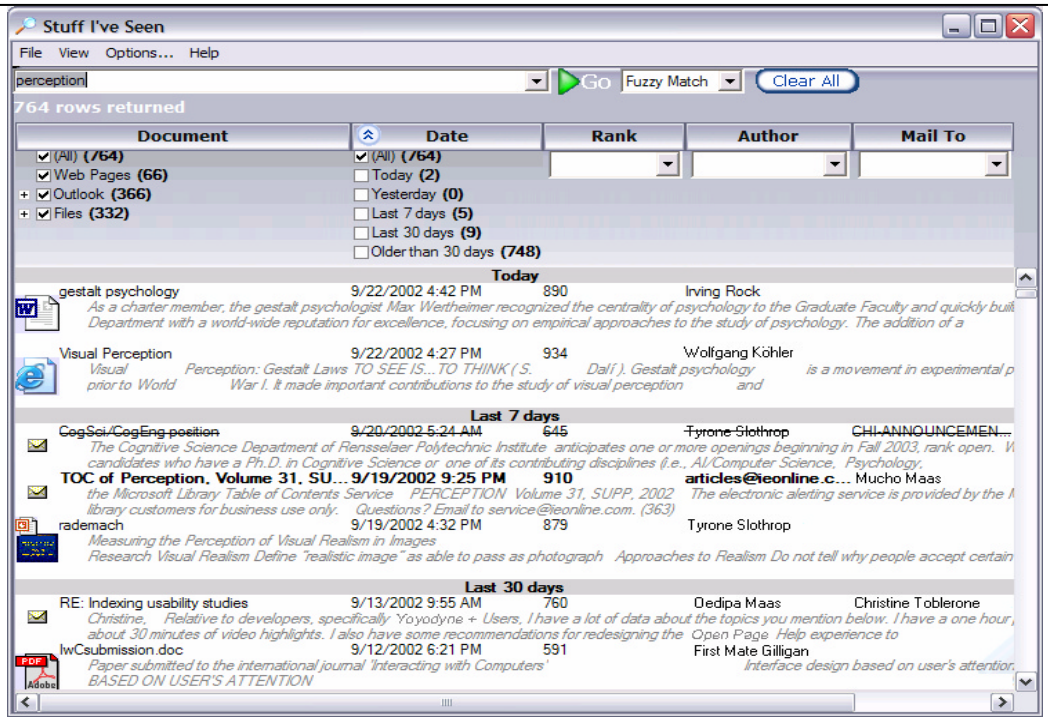

Figure 1. Screen shot of SIS interface, with the Top View.

morphological variants are also used (e.g., *car* matches *cars* as well), or fielded search in which matches are restricted to certain fields (e.g., author="Jane Doe").

The search results are shown in the lower portion of the display. In Figure 1, results include a preview showing the first 300 characters of a message as well as thumbnails for images and PowerPoint files. The previews can be turned off, increasing the number of results displayed. Five fields are present in the default view: Document Title, Date, Rank, Author and MailTo. Additional fields (File Type, Mail CC, Mail HasAttachment, Message Type, Message Read, Path, Size, Title) are available through an options menu. By default the results are sorted by either Date or Rank (different versions were deployed). Clicking on any column header sorts the results by that column. When Date is the sort field, markers showing the main date groupings (today, yesterday, etc.) are displayed to help group the results visually. The scroll bar on the right allows users to quickly move through the results.

Results lists can be further refined by selecting filters. In the Top View, filters for each column are located at the top of the column just below the column header. Checkboxes are shown when there are only a few alternatives (e.g., Document Type and Date), and text boxes with drop down lists are used when there are many possible alternatives (e.g., author). Filters can be applied even when the text box is empty. This enables users to find all items from a certain date range, all items from a specific person, etc.

The user interface shown in Figure 1 is somewhat complex and the filters at the top reduce the number of results that can be displayed. An alternative, called the Side View, is shown in Figure 2. This interface has the main query box and list view of results, but the filters have been simplified and moved to the side. In this view, filters are revealed serially. Selecting a specific item type, like Outlook, filters the results by that type and exposes additional fields that the user can specify for fielded search. In the case of Outlook items, the user can specify From, To, Path or whether the item has an attachment.

The Side View has the advantage that it's somewhat easier to understand and is less cluttered. And, because the filters are moved to the side, more results can be displayed. In contrast, the Top View is considerably more flexible. For instance, in the Side View it is not possible to filter by multiple types of items (e.g., Outlook *and* Web Pages), or by a specific column across all document types (e.g., author). In addition, in the Top View, the filters are associated with the columns they affect to create a more 'direct' filtering experience.

In both interfaces, double clicking on an item opens it up in the appropriate native application. Right-clicking an item brings up a context menu that allows the user to go to the folder containing the item (for mail and files), or to the cached version (for web pages).

SIS EVALUATION

We report qualitative and quantitative data from 234 people who used SIS during a six week period from August 1st, 2002 through September 16th, 2002. (The SIS project team was omitted from these analyses.) Our users come from variety of backgrounds, including consulting, legal, product support, administration, sales, and software development. Our user group includes individual contributors, managers, and executives.

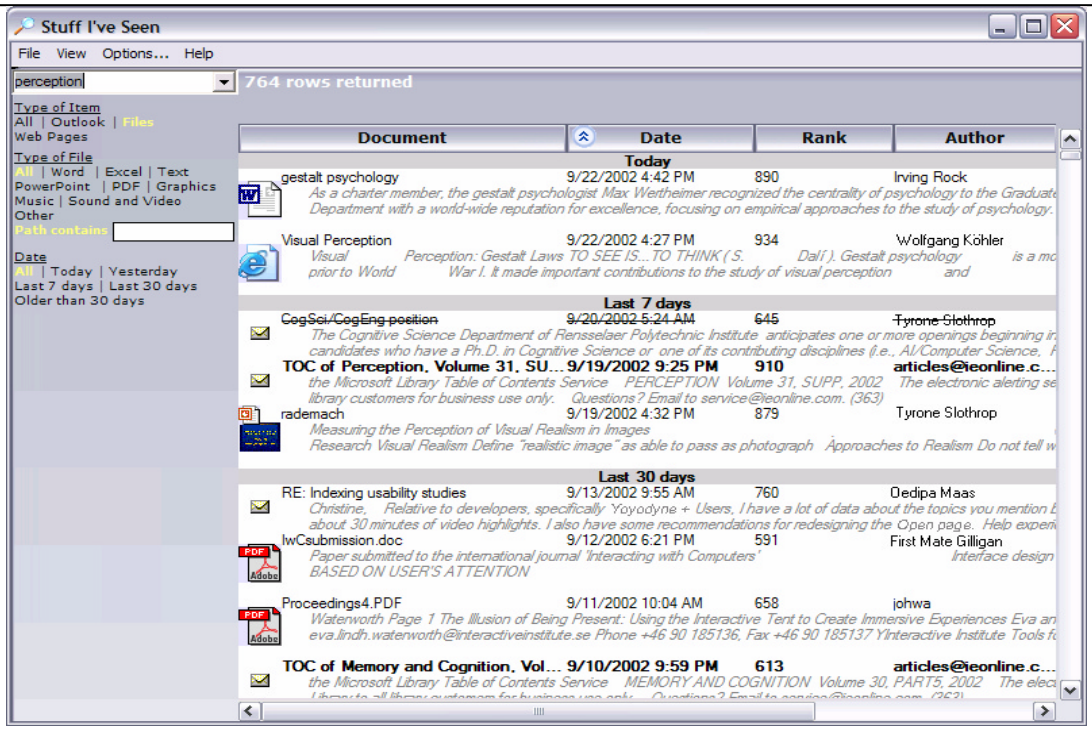

Figure 2. Screen shot of SIS interface, with the Side View.

We studied SIS using two main techniques: questionnaires and log file analysis. These techniques allow us to get a broad sampling of user activities in the context of natural use of the system. The questionnaires focused on how people organize information before and after SIS and about their experiences with SIS. The log data provided detailed information about the nature of user queries, interactions with the query interface, and about properties of the items retrieved. We have also conducted more focused and controlled laboratory experiments to inform the design of new visualization techniques [26].

Before people installed SIS, we asked them to take a brief survey about their current behavior in searching for e-mail, files, and web pages. We also asked people to categorize themselves as people who did not file e-mail, filed e-mail in bursts, or filed e-mail regularly (the same categories as described in [30]). Then, after about a month of use, a longer usage questionnaire was distributed to all SIS users.

In addition, SIS was instrumented to record all user actions with the interface. Examples of actions logged include: query text, use of filters, and the number of results returned by each query. For privacy reasons, we did not log any information about what content was indexed, or the content of search results. In addition to looking at overall interaction patterns, we also randomly deployed different versions of SIS to different users to explore differences between the Top and Side views and different default sort orders.

Log Data

During the time period studied, SIS users executed more than 8200 queries. They issued an average of 4.4 queries per day, but there was high variance with some users issuing no queries on some days and one user issuing 45 queries in one day. Users took advantage of the system on 84% of the work days they had it installed.

Query Characteristics

Although our users are arguably more computer savvy than the typical web user, many characteristics of their queries are similar to those reported in analyses of web query logs [27,28]. Only 7.5% of the queries involved explicit Boolean operators (AND, OR, NOT, +, -), phrases, or field restrictions specified in the main query box (e.g., from="Jane Doe"). Queries were short, averaging 1.59 words. This is somewhat shorter than the 2.16 reported in [28] or 2.35 in [27]. Short queries suffice in SIS because the local index and rich client allow users to quickly sort and filter results. In addition, personal content stores are smaller than the Web (ranging from 5k to more than 100k items for our users).

Although field restrictions were seldom used in the main query box, they were used frequently through direct manipulation of filters in the interface. Forty eight percent of the queries involved a filter specified using the checkboxes in the Top view (or selection in the Side view). The most common filter was to select one or more file types (selecting only email was the most frequent restriction, followed by selecting not email). The next most common filter was to restrict to a specified date range. Filters also account for many query refinements. Fifty percent of the query refinements involved filters, 35% involved changes to the query string, and the remainder involved changes to the display either by sorting or changing columns.

Given the work setting of our study and the personal nature of the stores, it is not surprising that the content of queries was different than that reported by Spink et al. [28] for

general web queries. The most common query types in our logs were People/places/things, Computers/internet and Health/science. In the People/places thing category, names were especially prevalent. Their importance is highlighted by the fact that 25% of the queries involved people's names suggesting that people are a powerful memory cue for personal content. In contrast, general informational queries are less prevalent.

Opening Items from the Search Results

There were over 8000 searches performed and nearly 2500 files opened using SIS. Several files could be opened after one search, and not all searches led to files being opened. The failure to open items after a search is difficult to interpret. It could mean that the search was a failure, or that the search results were used in other ways. The preview and metadata shown in the interface often provided the needed information. For items that were opened, we recorded the type, date and position in the list of results. Email was by far the most common type opened (76%), followed by web pages (14%) and files (10%). The most common file types were Microsoft Word (14%), plain text (11%), and Microsoft PowerPoint (11%), with the remaining types accounting for less than 10% each.

We also looked at the time distribution of opened items. Figure 3 shows the number of items opened as a function of time. Overall, 6.6% of the items opened were first seen that day, 21.9% within the last week, 45.9% within the last month, and 89.4% during the last year. Not surprisingly, recent items are accessed frequently, but the distribution has a long tail with items up to eight years old being opened. These long-tailed distributions have been reported for a variety of information access activities, but to our knowledge never before for personal items [6,7,25,32].

Figure 4 plots the access patterns on a log-log scale, and focuses on just recent items. The linear fit in the log scale is quite good (r^2 = .62, p << .0001). The fitted function is *log(Frequency) = -0.68*log(DaysSinceItemModified) + 2.02*. Others have reported similar power functions for re-access to web pages [9] and human memory [3]. The fitted parameters were similar for file, web and email content considered separately, except that email has a higher intercept (indicating more accesses) and a somewhat steeper slope (indicating a shorter effective life).

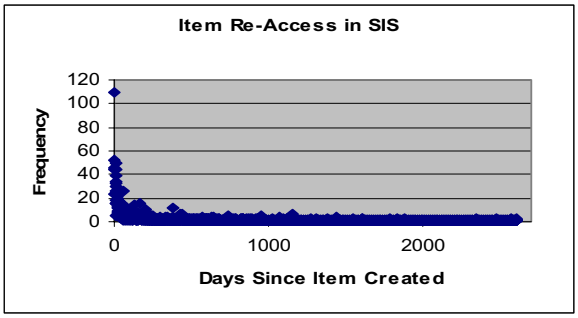

Figure 3. Frequency of access for items over time.

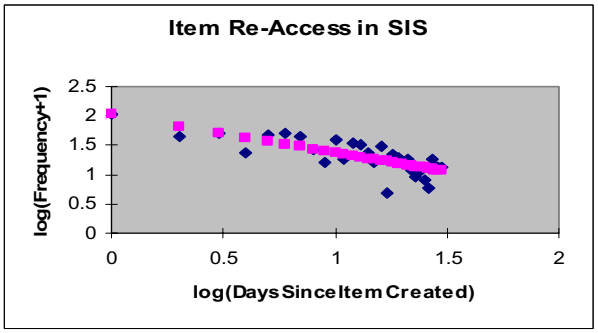

Figure 4. Frequency of access for items (log-log scale for 1 month as in [10]).

Interface Experiments

We randomly deployed different versions of the search interface to different users. Our logging data allow us to look at how many users changed these default settings.

Top vs. Side Layout

Half of the users started with the Top view and half with the Side view. (Because of logging problems we were unable to determine the starting layout for four participants.) Table 3 shows the total number of queries issued using the Top or Side view, broken down by which condition the user was initially assigned to by default. The usage data were analyzed with a mixed 2 (Starting UI, between subjects) x 2 (Used UI, within subjects) ANOVA. People who started with the Top view issued more queries than those who started with the Side view, but this difference was not reliable statistically ($F(1,228)=1.62$, $p=.21$). People who started with the Top view were less likely to switch to the Side view (34%) than vice versa (44%) and this effect was reliable ($t(228)=2.04$, $p<0.05$).

START:	**Top**	**Side**	**Total**
N	115	115	230
USE:			
Top	3043	1567	4610
Side	1621	1986	3607
Total	4664	3553	8217

Table 2. Use of layouts for results presentation.

Some of the more frequent switching to the Top view is likely due to the richer search interface which encouraged frequent filtering and allowed for fielded search. We are currently examining the details of query patterns to better understand this effect.

Rank vs. Date Sort

About half of the users started with results sorted by Date and the other half with results sorted by Rank. (Because of logging problems we were unable to determine the starting sort order for one participant.) Rank is determined by an Okapi-based algorithm, and Date is the last time an item was modified. Rank is the most common way to order search results in popular research and internet search engines. Date is a reasonable alternative for searching over personal content since people often remember roughly

when something happened. In addition to these two fields, users could sort on any of the other fields such as Author, Size, Path, Title, etc. Table 3 shows the number of queries sorted by various fields, broken down by which sort condition the user was assigned by default.

START:	Date	Rank	Total
N	111	122	233
USE:			
Date	3062	1975	5037
Rank	508	1530	2038
Title	250	186	436
Author	340	83	423
Path	73	93	166
To	57	52	109
Other	26	19	45
Total	4316	3938	8254

Table 3 Use of sorting options for results presentation.

The usage data were analyzed with a mixed 2 (Starting UI, between subjects) x 2 (Used UI, within subjects) ANOVA. People who started with Date sort issued somewhat more queries than those who started with Rank sort, but the difference was not reliable ($F(1,231)=0.24$, $p=0.60$). Regardless of which sort order people started with, they issued more queries in which they sorted the results by Date ($F(1,231)=22.8$, $p<<0.001$). The difference was reliable for those who started with Date ($t(110)=4.90$, $p<0.01$) but not for those who started with Rank ($t(121)=1.15$, $p=0.26$). The next most popular sort attributes after Date and Rank were Author and Title. The fact that users frequently switch to sorting by Date suggests that Date is a more useful attribute than Rank for finding personal items.

Questionnaire Data

In our questionnaires, we assessed how frequently and easily people searched for information both before and after using SIS. Forty five people responded to both questionnaires. Differences in the pre- and post- measures give us an indication of overall effect of the SIS system on user's searching behaviors. The questionnaires were completed about a month apart. Participants were asked to report on recent searching behavior using both Likert scale questions (e.g., *"When I need to search for a web page that I have seen before, it is easy for me to find it quickly."*) and frequency estimates (e.g., "*Yesterday, approximately how many times did you use a search engine to find a web site that you had previously visited?*"). We asked similar questions for search over email stores and the local file system as well. Figures 5 and 6 show the results of the ease of finding as measured by the Likert scale (Figure 5, top), and frequency of use estimates (Figure 5, bottom).

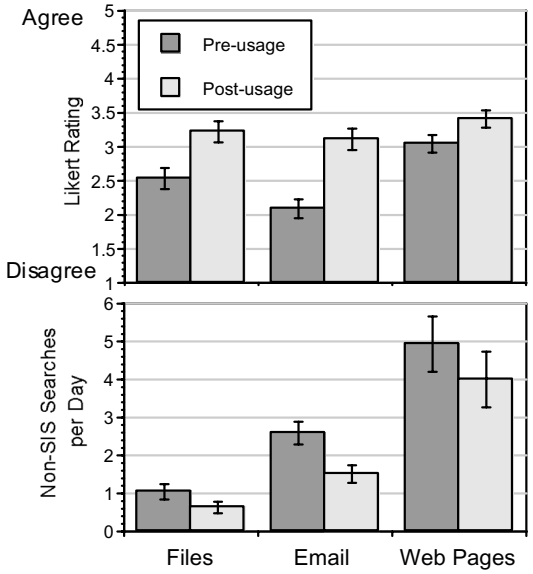

Figure 5. Ease of finding information, and frequency of use estimates, before and after SIS.

A 2 (Pre vs. Post Usage) by 3 (Content Type) repeated measures ANOVA was performed for the rating and frequency data. Users' ratings of ease of finding were higher after SIS than before ($F(1,43)=45.34$, $p<<0.001$). For all types of information seeking (files, email, web), users were more likely to say that they could find information quickly after using SIS (paired t-tests: $t(43)=2.39$, $p<0.01$ for files; $t(44)=6.64$, $p<0.01$ for email; $t(44)=4.35$, $p<0.01$ for web pages). There was also a reliable main effect of Type ($F(2,86)=13.47$, $p<0.001$) and an Interaction ($F(2,86)=6.76$, $p<0.002$). The interaction was due to the small difference for web pages, which isn't surprising since good web search tools exist. Larger differences were evident for email and files. With SIS, people found it equally easy to find all three kinds of information (light bars in Figure 5, top).

Estimates of how often they searched for information also changed after using SIS. Users estimated that they searched less frequently using native applications after installing SIS ($F(1,43)=12.09$, $p<0.001$), as shown in the bottom of Figure 5. There was a main effect of Type $F(2,86)=20.89$, $p<<0.001$) and no reliable Interaction $F(2,86)=0.74$, $p<0.48$. The differences were reliable for files ($t(44)=2.22$, $p<0.03$) and email ($t(44)=4.77$, $p<0.01$), but not for web ($t(44)=1.49$, $p<0.15$) although the difference was in the same direction. The decreased use of search in native applications was because they were using SIS instead.

The questionnaire also asked about participants' current filing strategies using the categories developed by Whittaker and Sidner [30]. Of the respondents, 39% were Frequent Filers (clean their inbox daily), 52% were Spring Cleaners (clean their inbox periodically), and 9% were No Filers (no use of folders). Good search over personal content decreases the need for maintaining complex file

and mail hierarchies, and we will explore changes in filing and search behaviors in a longitudinal study of SIS users.

The questionnaire also asked for general information about SIS. Some interesting items are listed in Table 4. These questions were answered on a Likert scale, where 1=strongly disagree and 5=strongly agree. Users were overwhelmingly positive about SIS, claiming that a SIS-like service should be an essential functionality for any computer system. They also very much liked the item previews, but they had trouble understanding different kinds of searches such as fuzzy match and Boolean queries. Finally, note that they reported SIS was used much more for email than for web pages, mirroring the log findings.

Question	Rating
SIS-like search service should be an essential functionality in any computer.	4.55
The previews of items that SIS provides are useful	4.06
I typically find things very easily with SIS	3.89
I typically use SIS to look for e-mail messages	3.87
I would get more value out of SIS if it indexed documents over all of my machines rather than just one client.	3.67
I would likely put less effort into maintaining a detailed set of folders for my files if I could depend on SIS to find what I am looking for.	3.45
I typically use SIS to look for web pages	3.00
It is easy to understand the different kinds of searches (e.g., exact, fuzzy, and / or) that I can make with SIS.	2.94
I use advanced query syntax to help me find what I'm looking for.	2.35

Table 4. Subset of post-installation questionnaire items, 1=strongly disagree, 5=strongly agree.

User Comments

When people filled out the SIS usage survey, we also gave them the opportunity to write general comments about SIS. The most common positive comments were about the speed of access, and the unification across different sources of information including archives. People said that SIS was often quite useful when they could only remember one or two vague things about an item, e.g, a general time frame or a rough topic. People also said that SIS was helpful for finding things that were 'buried' or filed in the wrong location. Even people who regularly file email and documents, often misfile information, or create new folders with slightly different names. SIS searches make it easier to find these items which become buried in folder-based access. The most requested new feature is unified access across multiple machines.

SUMMARY AND FUTURE WORK

We have designed, deployed, and evaluated a system that provides unified access to information a person has seen, regardless of where it came from. Initial findings are quite positive. They show that people find information easily using SIS, and use other search tools less frequently after installation. Filters such as date and type are frequently used to hone in on relevant items. Filters are easily specified in the interface and, coupled with fast client-side processing, encourage an iterative refinement strategy. Date and people names, in particular, provide rich contextual cues for retrieval, while standard ranking functions seem less important in the context of personal information.

We are continuing to develop the system and interface in several directions. One area for improvement is the overall performance of the search engine. Indexing should happen as quickly as possible without disturbing users' main activities, and we continue to tune parameters to accomplish this. Queries sometimes take longer than we would like, so we have added an explicit "go" button that is used to start queries (instead of having queries start automatically anytime filters are changed).

Another area for improvement is the user interface. The presentation of results is still a fairly standard list view. We are exploring different visual presentations of results including timeline visualizations with personal landmarks to further tap individuals' memories for their own content as well as summary views that collapse across individual items [26]. The presentation of faceted metadata developed by Hearst [31] and colleagues is also of interest.

We have developed infrastructure for users to tag their content with meta-data via SIS. Fast and effective search coupled with simple tagging could greatly reduce the need to maintain separate organizational structures for files, email and web pages. Others have also explored systems that move away from file hierarchies (e.g., Placeless System [11] and MyLifeBits [14]) and we will do so in the context of SIS with a rich user base.

Finally, we would like to extend the prototype to cover information on the fringes of awareness (i.e., Stuff I Should See) and to shared collaborative retrieval settings.

ACKNOWLEDGMENTS

We would like to thank Kyle Peltonen, Dmitriy Meyerzon and Eugene Samsonov for their help with the MS Search components, and all the SIS users for trying new software and providing ongoing feedback.

REFERENCES

1. Adar, E., Karger, D. and Stein, L. A. (1999). Haystack: Per-user information environments. *Proceedings of CIKM'99*, 413-422.
2. Ahlberg, C., Williamson, C. and Shneiderman, B. (1992). Dynamic queries for information exploration: An implementation and evaluation. *Proceedings of CHI'92*, 619-626.
3. Anderson, J. R. and Schooler, L. A. (1991). Reflections of the environment in memory. *Psychological Science, 10*, 396-408.
4. Barreau, D. and Nardi, B. A. (1995). Finding and reminding: File organization from the desktop. *SIGCHI Bulletin, 27(3)*, 329-339.
5. Belkin, N., Cool, C., Kelly, D., Lin, S-J., Park, S. Y., Perez-Carballo, J. and Sikora, C. (2001). Iterative

exploration, design and evaluation support for query reformulation in interactive information retrieval. *Information Processing and Management, 37(3)*, 403-434.

6. Broder, A., Kumar, R., Maghoul, F., Raghavan, P., Rajagopolan, S., Stata, R.. Tomkins, A. and Wiener, J. L. (2000). Graph structure in the Web. *Proceedings of the 9th International WWW Conference*, 309-320.
7. Burrell, Q. L. (1980). A simple stochastic model for library loans. *Journal of Documentation, 36(2),* 115-132.
8. Bush, V. (1945). As we may think. *Atlantic Monthly, 176*, 101-108.
9. Catledge, L. and Pitkow, J. (1995). Characterizing browsing strategies in the World Wide Web. *Computer Networks and ISDN Systems, 27(6),* 1065-1073.
10. Cockburn, A. and Greenberg, S. (1999). Issues of page representation and organization in web browsers revisitation tools. *Proceedings of OzCHI'99*, 7-14.
11. Dourish, P., Edwards, W. K., LaMarca, A. and Salisbury, M. (1999). Presto: An experimental architecture for fluid interactive document spaces. *ACM Transaction on Computer-Human Interaction, 6(2),* 133-161.
12. Erickson, T. (1996). The design and long-term use of a personal electronic notebook: A reflective analysis. *Proceedings of CHI'96*, 11-18.
13. Fertig, S., Freeman, E. and Gelernter, D. (1996). Lifestreams: An alternative to the desktop metaphor. *Proceedings CHI'96*, 410-411.
14. Gemmell, J., Bell, G., Lueder, R., Drucker, S. and Wong, C. (2002). MyLifeBits: Fulfilling the Memex vision. *Proceedings of ACM Multimedia'02*, 235-238.
15. Greenberg, S. (1993). *The Computer User as Toolsmith: The Use, Rreuse and Organization of Computer-Based Tools.* Cambridge, MA: Cambridge University Press.
16. Huynh, D., Karger, D. and Quan, D. (2002). Haystack: A platform for creating, organizing and visualizing information using RDF. Available at http://haystack.lcs.mit.edu/papers/computer-network2002.pdf.
17. Jones. W. P., Dumais, S. T. and Bruce, H. (2002). Once found, what next? A study of 'keeping' behaviors in the personal use of web information. *Proceedings of ASIST 2002,* 391-402.
18. Jones, S. R. and Thomas, P. J. (1997). Empirical assessment of individuals' 'personal information management systems'. *Behaviour and Information Technology, 16(3)*, 158-160.
19. Kaasten, S. and Greenberg, S. (2001). Integrating back, history and bookmarks in web browsers. *Proceedings of CHI'02,* 379-380.
20. Kaasten, S. and Greenberg, S. and Edwards, C. (2002) How people recognize previously seen WWW pages from titles, URLs and thumbnails. *Proceedings of Human Computer Interaction 2002,* 247-265.
21. Malone T. (1983). How do people organize their desks? Implications for the design of office information systems. *ACM Transactions on Office Information Systems, 1(1),* 99-112.
22. Marais, H. and Bharat, K. (1997). Supporting cooperative and personal surfing with a desktop assistant. *Proceedings of UIST 1997,* 129-138.
23. McKenzie, B. and Cockburn, A. (2001). An empirical analysis of web page revisitation. In *Proceedings of the 34th International Conference on System Science (HICSS34), CD Rom*.
24. Pitkow, J. and Pirolli, P. (1997). Life, death, and lawfulness on the electronic frontier. *Proceedings of CHI'97*, 383-390.
25. Recker, M. M. and Pitkow, J. (1994). Predicting document access in large, multimedia repositories. *Georgia Tech, Tech Report*, August 23, 1994.
26. Ringel, M., Cutrell, E., Dumais, S. and Horvitz, E. (2003). Milestones in time: The value of landmarks in retrieving information from personal stores. To appear in the *Proceedings of Interact 2003*.
27. Silverstein, C. Henzinger, M., Marais, H. and Moricz, M. (1998). Analysis of a very large Alta Vista query log. *SRC Technical Note 1998-014*, October 26, 1998.
28. Spink, A., Wolfram, D., Jansen, B. J., & Saracevic, T. (2001). Searching the web: The public and their queries. *Journal of the American Society for Information Science and Technology, 52(3*), 226-234.
29. Tauscher, L. and Greenberg, S. (1997). How people revisit Web pages: Empirical findings and implications for the design of history systems. *International Journal of Human Computer Studies*, *47(1)*, 97-138.
30. Whitaker, S. and Sidner, C. (1996). Email overload: Exploring personal information management of email. *Proceedings of CHI'96*, 276-283.
31. Yee, K-P., Swearingen, K., Li, K. and Hearst, M. (2003). Faceted metadata for image search and browsing. To appear in *CHI 2003*.
32. Zipf, G. K. (1949). *Human Behavior and the Principle of Least Effort*, Addison-Wesley.