BY EDWARD CUTRELL, SUSAN T. DUMAIS, AND JAIME TEEVAN

Searching to Eliminate Personal Information Management

Search systems can alleviate the need to organize personal information by helping us find it no matter where we encountered it, what we remember about it, and even if we forget it exists.

Delia wants to set up a lunch meeting at a restaurant her brother had recommended last week in an email message. She should be able to find the address using one of several organizational schemes she has developed to help her manage the vast quantity of paper and electronic information she receives every day. She knows she copied information about the restaurant into her address book, but she's not sure of the restaurant's name, making it difficult to look it up directly. She's not sure whether she filed the message in a folder relating to the main topic of her brother's email message or left

il faut
se méfier
des mots

it in her inbox. since she still needs to respond to him. She also remembers visiting the restaurant's Web page, so the information may be in her browser history.

Even though Delia is organized about managing her information, she doesn't always know exactly where to look for the information she knows she has encountered. She has tens of thousands of objects stored in multiple locations, including an address book, a calendar, a folder hierarchy for email, a different folder hierarchy for files, and yet another search and favorites mechanism for her Web history. Is all this organizational effort worth it?

Search engines are a familiar means of discovering new information, particularly on the Web. Search techniques can also be used to support access to a variety of personal information. In the physical world, a search like Delia's for a street address would depend on her having good organizational structures in place. But when information is stored electronically, rich search capabilities can augment or even replace explicit organizational structures as a means of locating and returning to information. Here, we explore the extent to which search can mitigate the need to organize one's personal electronic information. Organizational structures (such as learning, reminding, task management, and sense making) may support functions other than re-accessing information (see [7] for a discussion of the role of folders in project-related information).

Supplanting the need to organize personal information, search needs two key capabilities: First, it must cut across the many disparate sources of information we encounter daily; the address Delia wants could be in her address book, email folder, or browser history. And second, it cannot be restricted to keyword search but include other kinds of information associated with the item or context in which it was encountered (its metadata). Delia should be able to use whatever details she remembers to help her find the address of the restaurant; for example, she knows the email message was from her brother, the approximate time she received it, and that the restaurant's Web page included an image of a cornucopia and played an interesting musical theme.

Such rich associations characterize human memory and should be available in personal information management (PIM) systems to help people find information of interest.

SEARCHING PERSONAL INFORMATION

A search for personal information is different in many ways from a search in a vast unknown collection like the Web. Perhaps the most important difference is that people are familiar with many different characteristics of their information, as well as the context(s) in which they previously encountered them. Knowing many of them about what is being sought (including the fact that it exists) makes it all the more frustrating when we are unable to find it. Search capabilities that allow us to retrieve infor-
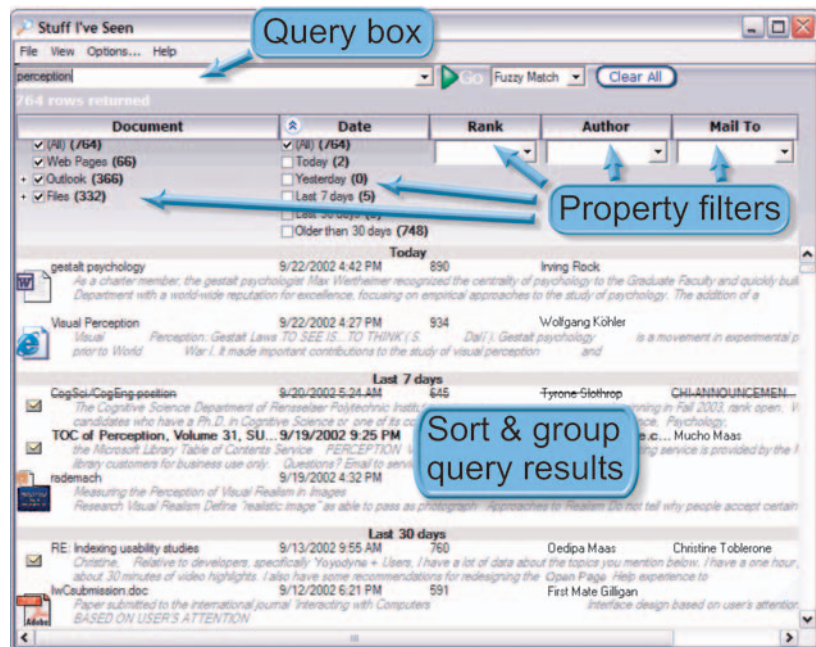
Figure 1. Screenshot of Stuff I've Seen interface. Users base their search on a variety of properties (such as date, file type, and author), as well as on keywords.

mation from a variety of sources, using a number of cues, in addition to keywords or folders, are critical for personal information access.

The idea of quickly and intuitively retrieving personal digital memories was popularized by Vannevar Bush, director of the U.S. Office of Scientific R&D during World War II, in his seminal article in 1945 [2]. Although today's technologies are very different from those Bush envisioned, today's desktop search tools fulfill many of his hopes. Here, we describe our experience developing and deploying one particular system at Microsoft Research—Stuff I've Seen (SIS) [6]. We have studied SIS extensively for the past several years using a variety of observational and experimental techniques and believe that our conclusions generalize to other similar systems. We note, however,

that a number of different desktop search tools have been developed over the past 30 years [4, 9] and that this functionality is now being built into the newest generation of operating systems for personal computers, including Apple Computer's Spotlight for Tiger OS X and Microsoft's Vista OS.

We developed SIS as a research prototype to provide unified access to electronic information a user would see regardless of how it was initially encountered (such as in email, files, calendar information, instant messages, Web pages, and digital photographs). Users do not need to do anything to explicitly store it. If they want to file an item in a folder, the information simply becomes an additional piece of metadata that can be used to assist retrieval. But SIS's rich search capabilities can be used whether or not an item is explicitly saved in a folder. People can search for information using any word associated with an item (analogous to Web search), as well as many different kinds of metadata or properties (such as what the item is, the item's dimensions, when it was encountered, and who created it). Figure 1 is a screenshot of the SIS user interface. At the top is a query box for specifying keywords or properties. Below are column headers for sorting by properties and other elements for filtering each of the properties. The search results are returned further down. The user interface enables keyword searching and property browsing to be tightly coupled through rich sorting, filtering, and grouping of results.

SIS was deployed as a voluntary download (still available) as a research prototype to Microsoft employees worldwide, many of whom still use it. They were representative of a variety of jobs typically found in large enterprises, including program management, sales, software development, administration, and executive management. We studied how they used it in their daily lives to access personal information. A descendant of SIS—Windows Desktop Search—is freely available from toolbar.msn.com.

INTERACTIVE AND ITERATIVE QUERIES

The queries from the SIS study participants were typically short (only 1.59 words on average, compared to 2.16 words reported on the Web [12]). Almost 50% of them were followed by iterations in which results were sorted and filtered in the SIS interface. This interaction allowed the participants to quickly refine their queries based on whatever contextual knowledge they could remember. Delia, for example, might enter the keyword "restaurant" into SIS, filter the results to show only email from her brother, sort the remaining results by date, and then scan for the email message from last week that

> Knowing many details about what is being sought (including the fact that it exists) makes it all the more frustrating when we are unable to find it.

contained the restaurant's address.

Searching in such an interactive and iterative fashion combines browsing and traditional keyword search. Any information need can be specified by whatever the searcher remembers: words in the content or metadata (such as sender, approximate time, or even folder name). One benefit of this iterative process is it allows users to recognize rather than recall what they're looking for. Users reported the system was particularly helpful when they remembered only vague attributes about the information they were looking for. The availability of many different attributes or access routes is a key SIS benefit compared to folder-based navigation that allows access using only a single attribute—the folder name.

PEOPLE AND TIME

Another way in which a SIS search for personal information differs from Web search is that the people associated with the information are an important retrieval cue. Indeed, over 25% of all queries issued in SIS included a person's name or email alias. This may be somewhat biased by the importance of email in a work environment, but it also reflects a more general characteristic of personal content. Personal information reflects the social milieu in which we organize our lives and memories. Delia, for example, knew that the email message she was looking for was from her brother Ben. People are a critical organizing element for personal information. While we are unlikely to know or care who created a Web page, the person who sent us an email message or who made a conference presentation is much better known to us and more relevant for retrieving a particular item.

Time is another important organizing feature in personal information. Although we are unlikely to know when a Web page was created or changed, we often remember roughly when we encountered the personal information we're looking for, especially relative to other events in our lives. Over 60% of all SIS search results were sorted by date. Other attributes (such as relevance, title, author, and folder) were also sometimes used to order results, but date was by far the most common. For Delia, sorting by date would

have allowed her to zero in on her email from last week, even though she did not remember the exact date.

The date attribute is especially notable because it highlights how what a user remembers about an item depends on context. Most items are associated with more than one date (such as when they were created, changed, and viewed). We found that the date users remember depends on the type of item they are look-
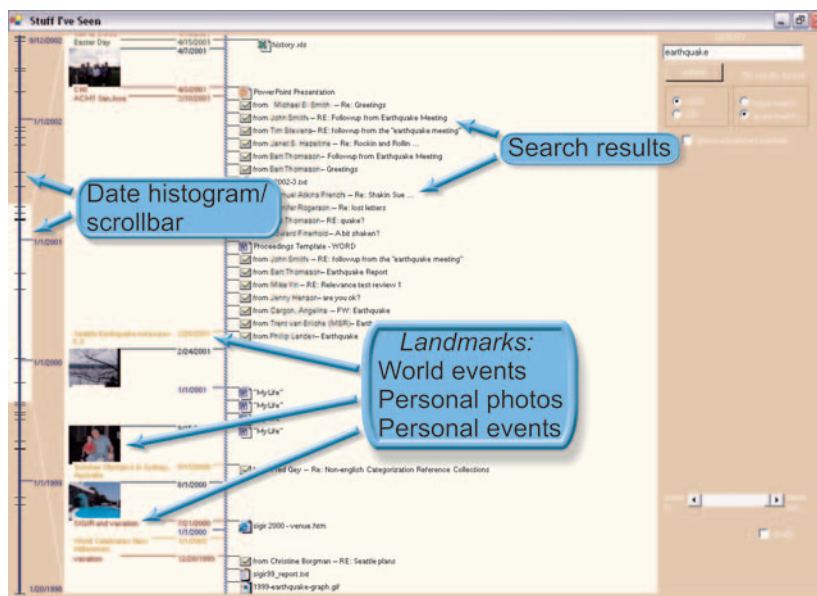
Figure 2. Screenshot of Stuff I've Seen Memory Landmarks interface. Search results are arranged with a timeline and events from a user's life (such as photos and calendar events) to provide memory scaffolding, or landmarks that help guide users to items of interest.

ing for. For example, for a calendar event, users typically remember when an appointment happened, not when the invitation was received or accepted. For Web pages, the memorable time is when a page was viewed; for photos, the date the photo was taken; and for email, when it was received. Therefore, the date shown in the SIS interface is an abstraction—the useful date—with different date information used for different types of items.

Time is such an important organizing feature for personal information that we developed a prototype timeline visualization (a standalone application using the same underlying index as SIS) as an alternative to the list view in SIS [11] (see [8] for another timeline interface and an extensive discussion of episodic retrieval). Research in cognitive psychology [3] has found that people remember information, particularly older information, not in terms of exact time, but in terms of key episodes, such as a child's birthday, exotic travel, and prominent world events like the attacks of 9/11 and the Indonesian tsunami of December 2004. Over 50% of the items users accessed through SIS were more than a month old, so it is important for SIS to support episodic access.

Figure 2 is a screenshot of the SIS Memory Landmarks interface providing a timeline presentation of search results, augmented with various landmark events. In the main section of the display, the results are ordered by time, much as they are in the SIS interface. The far left of the figure is a distribution of results over time, with the region in focus highlighted (December 1999–April 2001). The overview allows people to quickly identify time intervals of high search activity regarding the particular search topic. The landmarks section shows events that occurred at about the same time as the search results. These landmarks are used to identify time intervals of interest. Both public landmarks (such as holidays and key news events) and personal landmarks (such as important calendar appointments and digital photographs) provide anchors for access. A study of SIS users from Microsoft demonstrated that a landmark-enhanced timeline significantly improved user retrieval times and satisfaction levels when searching over their personal content [11]. The memory landmarks interface illustrates how search systems provide flexible access to personal information in ways that leverage the kinds of cues that people find memorable.

What are the implications for personal search, as our collections of digital information (and us users) age? With terabytes of personal information storage, how might search and retrieval work when we forget what we have? How can we search for something if we don't remember it exists?

FINDING WITHOUT SEARCHING

While personal search tools might eliminate much of what is currently considered PIM activity, search tools themselves may eventually be replaced with tools that proactively find information. People usually search for information in relation to ongoing tasks, and these task contexts can be used to support proactive information gathering. For example, as Delia responds to Ben's email message about the upcoming lunch meeting, that message can serve as a context for automatically finding related information (such as Ben's contact information, recent email from Ben, and other items related to the general topic of the message). All of this could be made available to Delia without her having to explicitly issue a query.

Some systems have also begun to take advantage of user context to proactively find task-relevant information [1, 5, 10]. They analyze the current context (such as an email message, a Web page, a television news story, or a current location), identifying important words or metadata and automatically generating queries to find related information. For example, the

One benefit of this iterative process is that it allows users to recognize rather than recall what they're looking for.

Implicit Query (IQ) prototype we developed (a standalone project using the underlying SIS index) [5] analyzes the email message the user is looking at and extracts important words from the body, subject, sender, and recipient fields. These words are automatically used in a query to the user's personal SIS index, and the results are shown as a side panel attached to the current message. We thought IQ would be helpful in sparing users the effort of generating queries, and indeed it is. But many people have also reported an unanticipated benefit of finding information, especially when they completely forgot they had anything related and would never have generated an explicit search on their own. In Delia's example, IQ could retrieve Web pages about an upcoming conference Delia and Ben are attending, reminding her to add the conference to the meeting agenda.

Many research challenges complicate development of systems capable of automatically finding information. Perhaps most important is designing an interface that balances awareness and distraction. To be useful, results must be visible and readily available, particularly when a user does not know relevant information is available somewhere out there. However, if the results are constantly changing based on what the user does, this can be distracting. A second challenge is how to deal with the complexity and opacity of implicitly generated queries. Unlike explicit search where the user specifies the parameters of the search, the relation between context and results returned can be complex and difficult to describe. A final challenge is how to support users when returning to known items in changing information landscapes, an interesting and important problem that's well beyond implicit query systems.

CONCLUSION

New search capabilities are changing the PIM landscape. Rich search capabilities make explicit filing and organizing far less important for retrieving personal information (though organization remains important for other reasons [7]). Several desktop search applications from Google, Microsoft, Yahoo, and other sources provide unified access to a range of personal information. Simple keyword search capabilities can be augmented with user interfaces to allow users to specify their information needs based on various cues (such as content, metadata, and task contexts) and to view and refine results quickly and flexibly.

Our experience with SIS and IQ indicates some of the ways search for personal content is different from other forms of search. Support for rich metadata (such as people, time, task contexts, and events) is critical for finding information users have previously encountered. But systems like those described here are just the beginning. In addition to explicit search, they will automatically provide information related to a person's task context. They will go beyond helping us find Stuff I've Seen and toward identifying Stuff I Should See. **c**

REFERENCES
1. Budzik, J., Hammond, K., and Birnbaum, L. Information access in context. *Knowledge-Based Systems 14,* 1–2 (Mar. 2001), 37–53.
2. Bush, V. As we may think. *Atlantic Monthly 176,* 1 (July 1945), 101–108.
3. Davies, G. and Thomson, D., Eds. *Memory in Context: Context in Memory.* John Wiley & Sons, Inc., Chichester, England, 1988.
4. Dourish, P., Edwards, W., LaMarca, A., and Salisbury, M. Presto: An experimental architecture for fluid interactive document spaces. *ACM Transactions on Computer-Human Interaction 6,* 2 (June 1999), 133–161.
5. Dumais, S., Cutrell, E., Sarin, R., and Horvitz, E. Implicit queries for contextualized search. In *Proceedings of the International Conference on Research and Development in Information Retrieval* (Sheffield, U.K., July 25–29). ACM Press, New York, 2004, 594.
6. Dumais, S., Cutrell, E., Cadiz, J., Jancke, G., Sarin, R., and Robbins D. Stuff I've Seen: A system for personal information retrieval and reuse. In *Proceedings of the International Conference on Research and Development in Information Retrieval* (Toronto, July 28–Aug. 1). ACM Press, New York, 2003, 72–79.
7. Jones, W., Phuwanartnurak, J., Gill, R., and Bruce, H. Don't take my folders away! Organizing personal information to get things done. In *Proceedings of the Conference on Human Factors and Computing Systems* (Portland, OR, Apr. 2–5). ACM Press, New York, 2005, 1505–1508.
8. Lansdale, M. and Edmongs, E. Using memory for events in the design of personal filing systems. *International Journal of Man-Machine Studies 36,* 1 (Jan. 1992), 97–126.
9. Quan, D., Bakshi, K., Huynh, D., and Karger, D. User interfaces for supporting multiple categorization. *In Proceedings of Interact 2003, the Ninth IFIP TC13 International Conference on HCI* (Zurich, Switzerland, Sept. 1–5). IOS Press, Amsterdam, The Netherlands, 2003, 228–235.
10. Rhodes, B. and Maes, P. Just-in-time information retrieval. *IBM Systems Journal 39,* 3–4 (July 2000), 685–704.
11. Ringel, M., Cutrell, E., Dumais, S., and Horvitz, E. Milestones in time: The value of landmarks in retrieving information from personal stores. In *Proceedings of Interact 2003, the Ninth IFIP TC13 International Conference on HCI* (Zurich, Switzerland, Sept. 1–5). IOS Press, Amsterdam, The Netherlands, 2003, 184–191.
12. Spink, A., Wolfram, D., Jansen, B., and Saracevic, T. Searching the Web: The public and their queries. *Journal of the American Society for Information Science and Technology 52,* 3 (Feb. 2001), 226–234.

EDWARD CUTRELL (cutrell@microsoft.com) is a researcher in the Adaptive Systems and Interaction group at Microsoft Research, Redmond, WA.
SUSAN T. DUMAIS (sdumais@microsoft.com) is a senior researcher in the Adaptive Systems and Interaction group at Microsoft Research, Redmond, WA.
JAIME TEEVAN (teevan@csail.mit.edu) is a Ph.D. candidate in the Computer Science and Artificial Intelligence Laboratory at the Massachusetts Institute of Technology, Cambridge, MA.