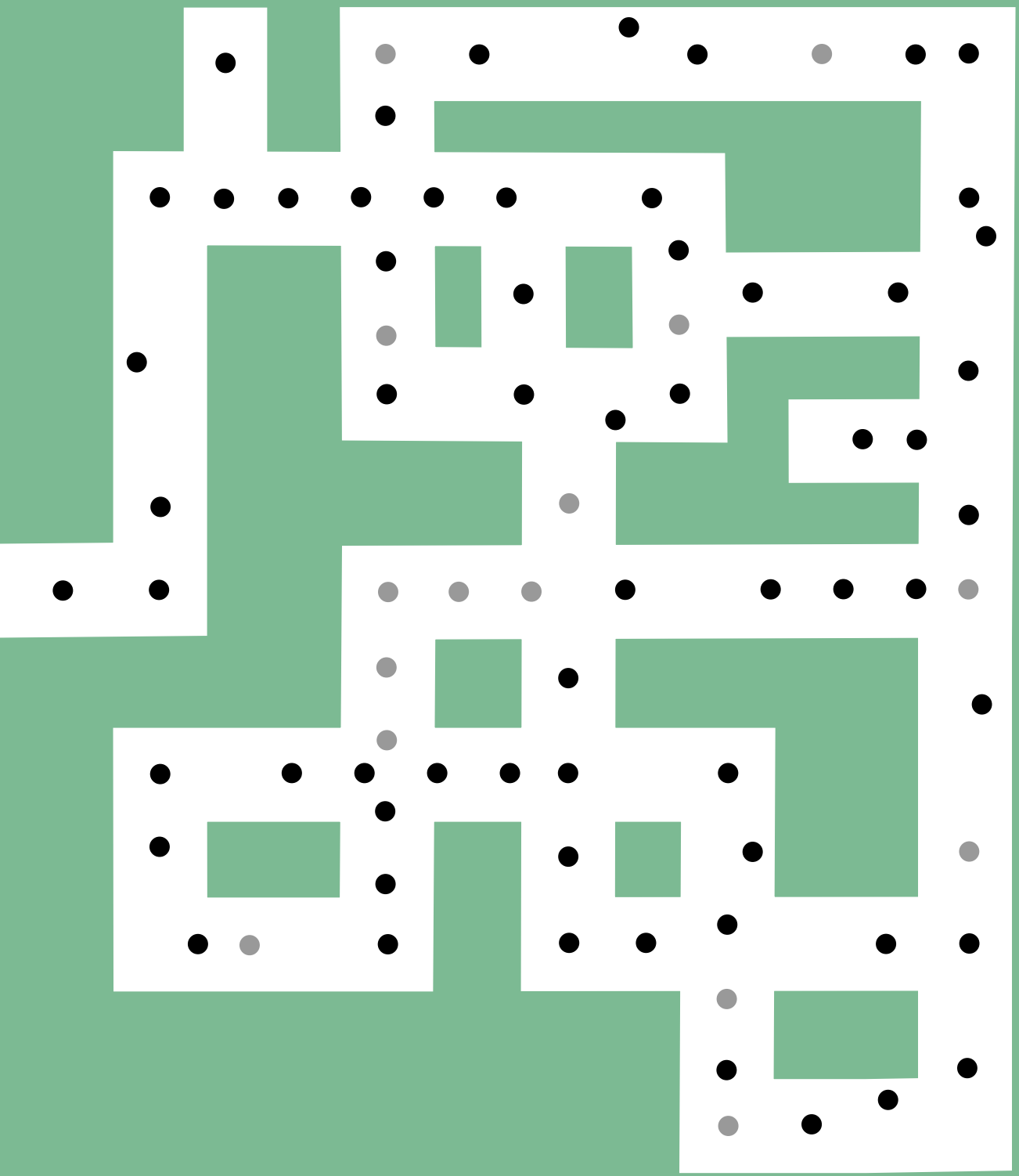
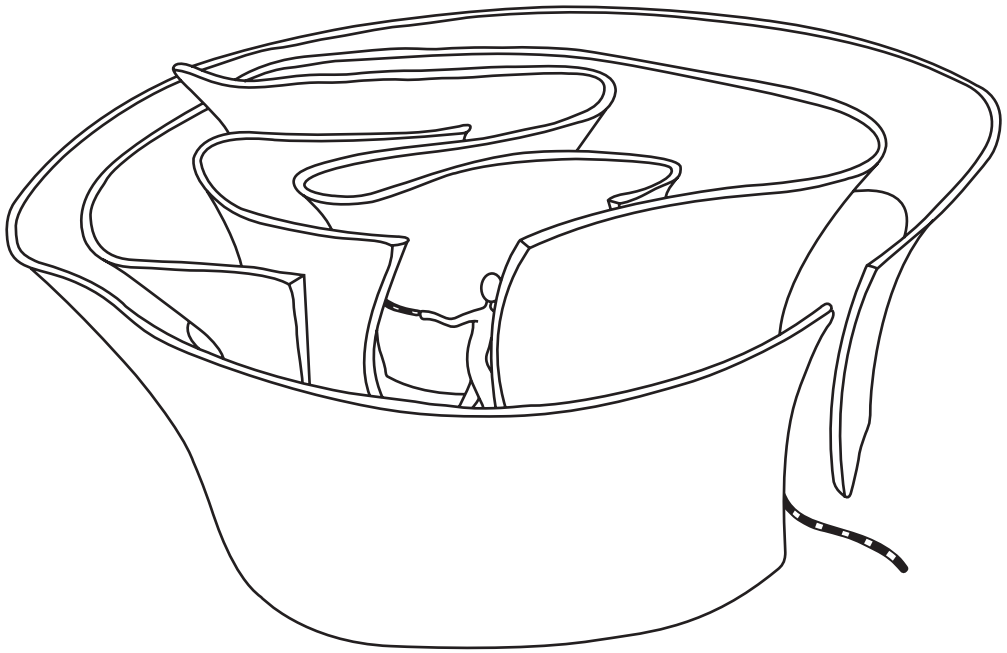
Chapter Four

Finding and re-finding: From need to information

Finding is multistep. Recall something about the information. Recognize it when we see it. Repeat as needed. But also, we have to remember to find it in the first place. Finding is not just about big acts to seek new information "out there." The many small acts to find the familiar can kill a day. *Wayfinding* treats finding as a journey—from a need and the situation prompting it, to information, and then back again.





I do not seek. I find.

Spanish painter and sculptor Pablo Picasso (1881–1973)

4.1 Starting out

"Where is it? I know it's here somewhere. Where on earth did I put it?" "It" can be the car keys or a pair of shoes. "It" can be an information item such as an email message or a tax-related document. Certainly some of the more painful, memorable failures of PIM relate to failures to find information that "I know is here . . . somewhere." In my PIM seminars, people report efforts to find information—especially paper documents such as a title to an automobile they wish to sell or a birth certificate or a passport—that may extend over a period of a week or more. Finding over such a period of time may be a collaborative effort[1] involving various members of a household or an office work team. Failures to find can be a real source of discord as frustration mounts and accusations and recriminations are exchanged ("What did *you* do with it?").

If finding activities, with their focus on the location, output, and use of information, represent an endpoint in PIM, their study is also a natural place to begin taking a closer look at PIM.

[1] For more discussion of finding as a collaborative, group activity, see Berlin et al. (1993) and Fidel et al. (2000).

The finding stage makes apparent, sometimes painfully so, larger failings in a person's practice of PIM.

We are aware of larger acts of finding especially when these involve significant time and creative effort to complete or when these involve a conscious search as supported, for example, by a web service or a desktop facility. But large portions of a day can be consumed in many smaller acts of finding—the many look-ups required to fill out a form, for example, or the repeated references to a calendar in an effort to schedule a meeting. Though the many smaller acts of finding may each take only a little time, together they can still add up to a significant portion of a day.

In this chapter, we will see that finding is as much about interaction as about end result. There are two senses in which this is true. First, we can find the targeted information and yet still feel a sense of failure unless the process of finding the information was reasonably short, pleasurable, and trouble-free. Second, even though our focus is on the targeted information, we may gain considerable benefit from the incidental interactions with information along the way to this targeted information. To take a simple example, if a person checks her calendar to see when today's staff meeting will take place, she may happen to notice that another meeting is also scheduled for later today. More generally, the path taken to targeted information can be a source of serendipitous discovery. Finding is about the journey as well as the destination.

These topics are explored in this chapter as we move through the following sections:

- *Getting oriented.* Research on finding, as a PIM activity, is placed in the context of a larger field of research on *information seeking*. The challenges of finding and the opportunities for tool support vary according to whether we've experienced the items we seek before and whether these are in a store that we own and that is (nominally at least) under our control. Focus in this chapter is primarily on efforts to find (re-find) information we've experienced before and that is inside our PSI in a store that we control. This section also considers the essential movement that underlies finding activities: from a current need to the access and use of information that meets this need.
- *Everyday finding: Death by a thousand look-ups.* We consider the many small acts of finding that can add up to a significant proportion of a day's time—and its frustrations.
- *Finding is multistep*. Any act of finding involves an interplay between *recall* and *recognition*. We recall something about the item we are trying to find that might help to narrow the scope of a scan to recognize the item (e.g., in a folder or search results listing). Also, the finding process overall must often *repeat* several times so that a complete set of items is assembled. Finally, we must *remember* to find in the first place. Finding can fail because of the failure of any of these steps.
- *The limitations in ideal dialogs of finding.* We consider an ideal finding dialog with the computer (i.e., with a computer-based search tool) to be one that is much like the dialog we might expect to have with a well-trained human assistant. The computer can make use of anything we can recall. The computer orders and represents candidate items to aid us in

our recognition of the item we seek. Support of a dialog like that between two people has many advantages but also one fundamental limitation: we may not always know or be able to express the things we need to find.

- *Wayfinding through the PSI.* We explore *wayfinding* as an alternate, complementary metaphor to finding as a dialog. The wayfinding metaphor gives emphasis to finding as a journey through the PSI—from a need and the situation prompting this need to information and then back again. As with any journey, a journey through the PSI can be serendipitous, yielding useful information we did not expect and would not have thought to ask for.

4.2 Getting oriented

What is information finding? We begin with Wilson's definition for information seeking (2000):

> The purposive seeking for information as a consequence of a need to satisfy some goal. In the course of seeking, the individual may interact with manual information systems (such as a newspaper or a library), or with computer-based systems (such as the World Wide Web). (p. 49)

This definition emphasizes the teleological or purposeful nature of finding: finding to satisfy a goal. Certainly goals matter. Failure to reach the goal, failure to find the sought-for information, is frustrating, costly, and memorable. For example, Sellen and Harper (2002) review studies suggesting that the average manager spends three hours a week looking for documents that have been misfiled.

But in a large number of everyday instances of finding, the end—if defined as eventually getting the information—is not in doubt. The information will be found one way or another, sooner or later. The bigger questions are how, and how long it will take. Can we find information easily and in the natural course of our efforts to get things done? Or is finding a separate, time-consuming, disruptive experience that takes us away from the other activities of a day?

As a supplement to Wilson's definition, this chapter takes the following slant on the finding activities of PIM:

> Information finding is an ongoing, minute-by-minute interaction with a large and growing PSI involving not only the needed information but also the information, organizing constructs, and tool support that are encountered and used along the way to this needed information and then back again to the situation prompting the need for this information. Needed information is found following paths through the PSI. Found information and the process of finding extend and further integrate the PSI.

Certainly the efforts of people to re-find information that they have previously saved involve interactions with the PSI. But a person's PSI is also invariably involved in efforts to find new information from a public space of information. The need behind a search on the Web for hotels in Montreal, for example, may be triggered by the glance at a personal calendar indicating that the trip to Montreal is only a month away. The search itself may be made easier, if ever so slightly, by having a favorite search service referenced on the home page—a part of the PSI. Or the web page may be quickly accessed through a web address that is specified

by a keystroke or two and completed by an "auto-complete" suggestion of the browser. The information the browser uses to complete the address is also part of the PSI.

Information, once found, is often applied to an item in the PSI such as an email message or a document being created. Also, the information itself or a path to the same may be kept in the PSI so that it can stay found—easier and more likely to be re-accessed again later. The "marks" (impacts) of information found may eventually be on, or in, the person[2] in the form of new understanding or new internal knowledge structures. In the meantime, however, people can at least hope that the marks of information found are in their PSIs.

4.2.1 Finding vs. re-finding in public stores vs. private stores

Activities of information finding can be placed in one of four quadrants in Table 4.1 according to two senses of "personal" as described in Chapter 2:

- Is the information controlled ("owned") by us? Most of us control (in principle) the information on the hard drive of our personal computer. We can move or delete it (though many times we're properly reluctant to do so for fear of creating problems). Most of us control very little if any information on the Web. We can't move or delete it.
- Have we experienced (seen) the information before?

Activities of finding in each quadrant are important to a practice of PIM. Each quadrant presents its own challenges and its own opportunities for tool support.

A. Re-finding information we control (and have seen before)

Getting back to "my" information is especially important in PIM. We took the trouble to keep it and sometimes to create it in the first place. We're not done with it yet and may never be. We may have partially read an email message, for example, or a document, with the intention to read more thoroughly later. Some information—articles, contacts, or a spreadsheet with passwords—may be general reference to which we return repeatedly. Desktop search facilities can help, especially if these are integrative in their ability to search across multiple information

Table 4.1 An event of finding can vary according to where targeted information is and whether the person is trying to find this information again

The information is . . .	Controlled by us	Not controlled by us
Seen before by us	A	B
Not seen before by us	D	C

[2] Kidd (1994) takes an extreme position that knowledge workers rarely need to revisit an information item once they have been "informed" by its contents.

forms, and even more so as these tools are increasingly integrated into our use of other tools.[3] Generalized support for automatic completions and fill-ins can also help in the reuse of smaller pieces of information such as email addresses, phone numbers, and account numbers (see the "What next for tool development" sidebar). Also relevant are efforts to automate the tagging and grouping of information items by task and efforts to situate information items in the context of the planning of a larger project.

B. Re-finding information on the Web

We also return to information on the Web. In a study covering six weeks of web visits for 23 participants, Tauscher and Greenberg (1997a, 1997b) found that there was a 58 percent likelihood that the next page seen by a person was a page the person had already accessed at some point in the past. Jones et al. (2003) found that people use a variety of different methods and supporting tools for returning to Web information. Especially popular are "do-nothing" methods that require no keeping forethought. These methods include (1) clicking through hyperlinks from a familiar starting point such as a web portal or home page, (2) searching again, and (3) "auto-complete" facilities that suggest completions to a partially typed web address where completions are drawn from web addresses for pages previously visited.

C. Finding new information (not seen before) on the Web and other public stores

There is an extensive body of work on information seeking and information retrieval that applies especially to events in the C quadrant.[4] It is important here to note that there is a strong personal component even in efforts to find new information, never before experienced, from a public store such as the Web. For example, our efforts to find information may be directed by an outline or a to-do list that we maintain in our personal space of information. Access to new information items may be through a query that we maintain in our personal space as a bookmark or even as a list of words we keep in written form (or "in mind"). Much more can be done to use existing personal information in efforts to find new information from a public space.[5] Equally important, much more can be done to situate our searches on the Web with respect to informational situations in the PSI that prompts these searches—a topic of further exploration throughout this book.

D. Finding new information that we control

The amount of information we (ostensibly) control continues to increase along with increases in the capacities of storage devices we own. When we use a desktop search facility, we may be

[3] For discussions on the ongoing integration of search functionality, see Cutrell, Dumais, and Teevan (2006) and Russell and Lawrence (2007).

[4] See, for example, Marchionini (1995), Marchionini and Komlodi (1998), and Rouse and Rouse (1984).

[5] For a discussion of possible uses of personal information in support of search, see Teevan (2006).

surprised by what we find—by what we "have" already. One challenge in tool support, discussed later in this chapter, is to call a person's attention to information he or she has already and that may be relevant to the current situation—and to do this without becoming a nuisance.

Note that two other chapters in the book stand in different, complementary relationships to the current chapter. Problems experienced during finding often originate as earlier failures of keeping and organizing, as explored in Chapter 5. Searching technology, as explored in Chapter 11, can support finding in ways less dependent on careful prior keeping and organization of information.

What are the differences, really?

How much do differences in where and how we find information matter as long as we get the information we need? Certainly there's a difference between new finding and re-finding. If people have a specific item in mind, their search is more focused. They have memories from previous encounters that they can use (or ought to be able to use) in order to narrow the scope of the current search.

What about differences between finding information we control vs. finding the information "out there" on the Web and other public repositories? We may increasingly have the experience of finding new information inside our PSI in a store that we control. Such information, though newly found, is often information we ought to have experienced (that is, known was there to be found) or might want to have experienced, even if we haven't—so far. Our reaction, for example, to the discovery of an unread email message sent to us by a friend a year or two ago—even if only an "fyi" pointer to Web site of possible interest—is likely different from our reaction had we discovered a pointer to the same web site in someone's blog instead. The email message is directed to us personally.

Most important, the experience of failing to re-find an information item is different when the targeted item resides in a store we control vs. the Web or some other public store. We're less surprised when a Web page visited yesterday is not available today. An access failure could be for any number of reasons beyond our control—frustrating, to be sure, but "these things happen."

Do we generally show the same kind of equanimity when information under our control (nominally at least) can't be re-found? Leave aside documents we have authored which may represent many hours of our own work. Consider, instead, an article written by someone else which could just as easily be found on the Web as on a local hard drive. The failure to find this article inside the PSI often stands for something much larger than the loss of the information in the article itself. Failure can come to represent a larger failure in our lives . . . a loss of control. "Where on earth is it? Am I losing my mind, too?" We take it personally.

This chapter's discussion will focus primarily on finding in quadrant A—where the effort is to find (re-find) information we've already experienced from a store that is under our control.

4.2.2 From need to information (and back again)

We're almost done with this section's brief orientation. The remaining task is to consider two points in relation to this chapter's subtitle.

1. The journey from need to information is round-trip. Once information is found, it is either applied—"used"—in the situation that prompted its retrieval, or it is possibly kept for use later. Activities of finding and re-finding need to be situated in this larger context.
2. Our needs change with every step we take. Our own understanding of a current need, as reflected, for example, in our descriptions or in our seeking behavior, is constantly changing. Many changes in our understanding of a need are brought about by the actions of finding (seeking) itself and by the information retrieved.

The journey from need to information is round-trip

Information management and use are interwoven. This central theme of the book encourages us to think beyond just the location and access of information. Information revealed through browsing or "web surfing" or as referenced in a results listing returned by a search query is subject to several different keeping decisions. Is this information useful? Now, for a current need, or later for an anticipated need? If later, does anything need to be done now to ensure its availability later on? Do reminders need to be set? Should the information or a pointer to this information be kept?

But considerations of the return trip apply even if information is used immediately and then discarded. How is the information used? Is the information sent out in an email message or used in a document? Is information used as is, or are steps taken first to interpret, make sense of, and integrate the information into a larger document?

We might dream of a high-fidelity Copy & paste (or Drag & drop) in which the information found is copied into a new information item without loss of formatting. Better, the reference to this information is also copied. Portions of this reference can automatically be included, for example, in a document's bibliography. Other portions of the reference might be included in a hyperlink that makes it easy to get back to the source for more information as needed and possibly with the excerpted information highlighted in context. In some cases, we might even want to subscribe to updates in the content of web sites from which information is excerpted.

Why stop here? Why not also situate an act of finding itself with respect to an informational context. For example, we see a line item in a spreadsheet budget and send an email to our group's financial person for clarification—an act of finding. Why not record this act on an optional overlay to the budget's display? Later, when we want to review the response to our inquiry, we then have two ways to return to the email we sent and its responses: either go back to the context (e.g., the spreadsheet budget) that prompted us to send the email, or try to access email responses from the context-free jumble of the inbox. Which would you pick?

Or, as another example, we see a web site describing a new product and are moved to search for blogs giving commentary on the tool. Rather than carefully saving useful results in a separate document or through a separate bookmarking facility, why not save them as an overlay to the web site that prompted us to search?

Alas, current support for the "return trip"—from information found back to the situation that prompted its finding—still falls far short of what we might hope for. Finding tools such as search facilities and email applications still function more as worlds unto themselves rather than as an integral part of our informational context. As we use these tools to access needed information, we're still mostly on our own in our efforts to return to the informational context prompting the finding in the first place. And, as Karger (2007) notes, the transfer of information from the source back to the current situation through Copy & paste (Drag & link) is often still text-only. Or worse, we get information in paper form or as a scanned image and must then either transcribe or attempt to use character recognition software.

The larger point here is that we need to think of finding activities as part of a larger journey: from need to information and then back again, to need and the use of information to meet this need. Doing so raises practical questions that are easily overlooked if the focus is only on finding the information itself.

Needs change with every step we take

Even our effort to describe a need creates a refinement in our understanding of it and, sometimes, leads us to abandon the need altogether. "I want to see information on Paris's nicest hotels. . . . On second thought, maybe I'd rather stay at a little pension in a quiet neighborhood."

Belkin (1993) notes, "There is by now a substantial literature, from both theoretical and empirical perspectives, on the non-specifiability of information 'needs'" (p. 59). Several models of information seeking move beyond what Belkin refers to as "the standard view of IR" (information retrieval) in which a person approaches an information system with a well-specified query and "the major issues of concern" are "the representation of texts, and of queries, and techniques for the comparison of text and query representation" (p. 56). Interactions with an information system can be characterized as a process of negotiation,[6] a dialog,[7] or a process in which a person's knowledge changes through interactions with the information retrieved, leading, in turn, to a reassessment of information need.[8]

Especially evocative is M. J. Bates's *berry-picking model* of search (1989), depicted in Figure 4.1, to account for situations in which search interest and the expression of this interest evolve over time as a function of results returned by previous searches. For example, a search for "hotels in Paris" may return one result for a bed and breakfast inn in Paris. The person might then have the thought, "Hmmm, maybe I would like staying at a bed and breakfast better than at a large hotel," which is then reflected in a follow-up search for "bed and breakfast inns, Paris." The new query reflects a shift in interest—affected by previously returned results but not simply an attempt to search within these results.

[6] See, for example, Taylor (1968).

[7] See, for example, Oddy (1977).

[8] See, for example, Belkin, Oddy, and Brooks (1982) and Belkin, Seeger, and Wersig (1983).

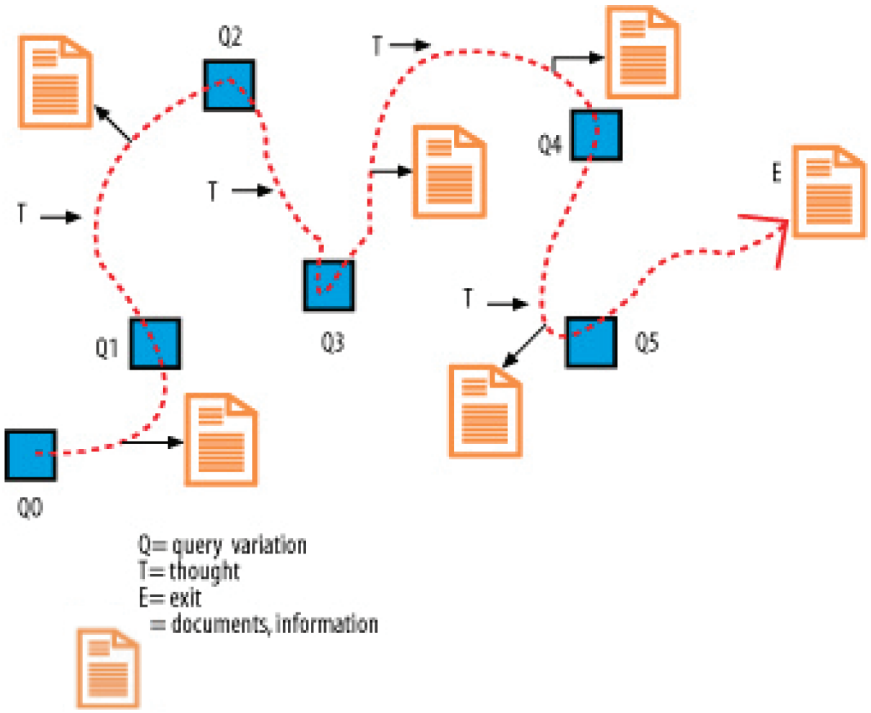


Figure 4.1 **The berry-picking model describes a situation in which a person's search query (Q*n*) wanders as a function of results returned from previous queries.** ***Source:*** **From *http://www.flickr.com/photos/morville/84894767/in/set-1812650/.***

The concept of information need is itself subject to many interpretations. In the dialog between student and teacher, for example, who has a better understanding of the student's information needs with respect to an assigned term paper? Taking a behaviorist's position, Wilson (1981, 2005) suggests abandoning the concept of "information need" altogether in favor of only focusing on observable information-seeking behavior. Dervin (1992) refers instead to "sense-making" activities as motivated by a person's perception of a gap in his or her understanding of a situation.

However, it is difficult to assess the effectiveness of various finding activities or the tools designed to support them without making some assumptions concerning the motivation behind and desired outcomes of these activities.

Rather than abandoning the notion of need, it may be better to acknowledge that the assessment of need changes radically by person and in the same person with every step taken. The student now perceives only a need to do the bare minimum in research required to get a passing grade so that he can have more time for soccer practice. He may perceive a different need in 10 years when he is looking for a job. Even now he may be persuaded by the teacher to do more work, or maybe he will continue researching just for the fun of learning more about the term paper topic.

No matter what the need (or motivation), even if the student does the bare minimum, he will face the same basic problems of finding. Moreover, problems of finding arise in many everyday situations having little to do with canonical information-seeking situations where the search is for information in a library or on the Web. It's time to talk about everyday finding.

4.3 Everyday finding: Death by a thousand look-ups

Nancy is filling out an expense reimbursement form (Figure 4.2) for her recent trip to Montreal to attend a trade show. The form takes her more than an hour to fill out rather than the allotted half hour she'd hoped would suffice. No wonder she's been postponing its completion (even though she could use the reimbursement).

Why the extra time? The mechanics of finding just take time. Each space on the form seems to force another effort to find the required information. Nancy's internal dialog with herself, as she works to complete the form, might go something like this: "Departmental code? Note sure, better look it up. Same for budget number. I remember the destination all right but better check ticket to be sure I get the date/time of departure and return correct. And then check again for cost of ticket. Same thing for car rental. Where is that receipt anyway? I drove to and from the airport. What's the distance to the airport again? I think I wrote it down somewhere. . . . Or better to look at a previous expense report." By the time Nancy has completed the form, she is frustrated by the time it has taken, and she is even more inclined to postpone the completion of similar forms in the future.

Or consider a Web page form (Figure 4.3) that Nancy filled out much earlier in order to reserve a hotel room for the business trip. Nancy needed to decide the dates to reserve, which in turn meant that she needed to find information concerning the trade show she wanted to attend, its key events, and other events in her work life and personal life that she wanted to make room for before and after the trip. She wanted to spend enough time at the trade show, but she was thinking that she might extend the trip if she could talk her spouse, Jim, into coming with her.

*** PLEASE INCLUDE ITINERARY AND ALL RECEIPTS ***

Name:	**Date:**
Departmental Code:	**Budget #:**
Destination:	
Purpose:	
Date/Time Departed:	**Date/Time Returned:**
Cost of Airfare or Train:	**Car Rental:**

Date	Hotel Room Rate(per night)	Meals(per day)	Taxi	Parking	Miles Driven	Conference or Registration Fees	Misc.	

Figure 4.2 **Finding activities account for much of the tedium in form-filling.**

In that case, she would have needed to specify "2 guests per room" rather than one. But first she needed to call Jim (another act of finding) to see if he might be interested in coming along (he was but couldn't after he checked his own schedule—more finding). Nancy needed to check as

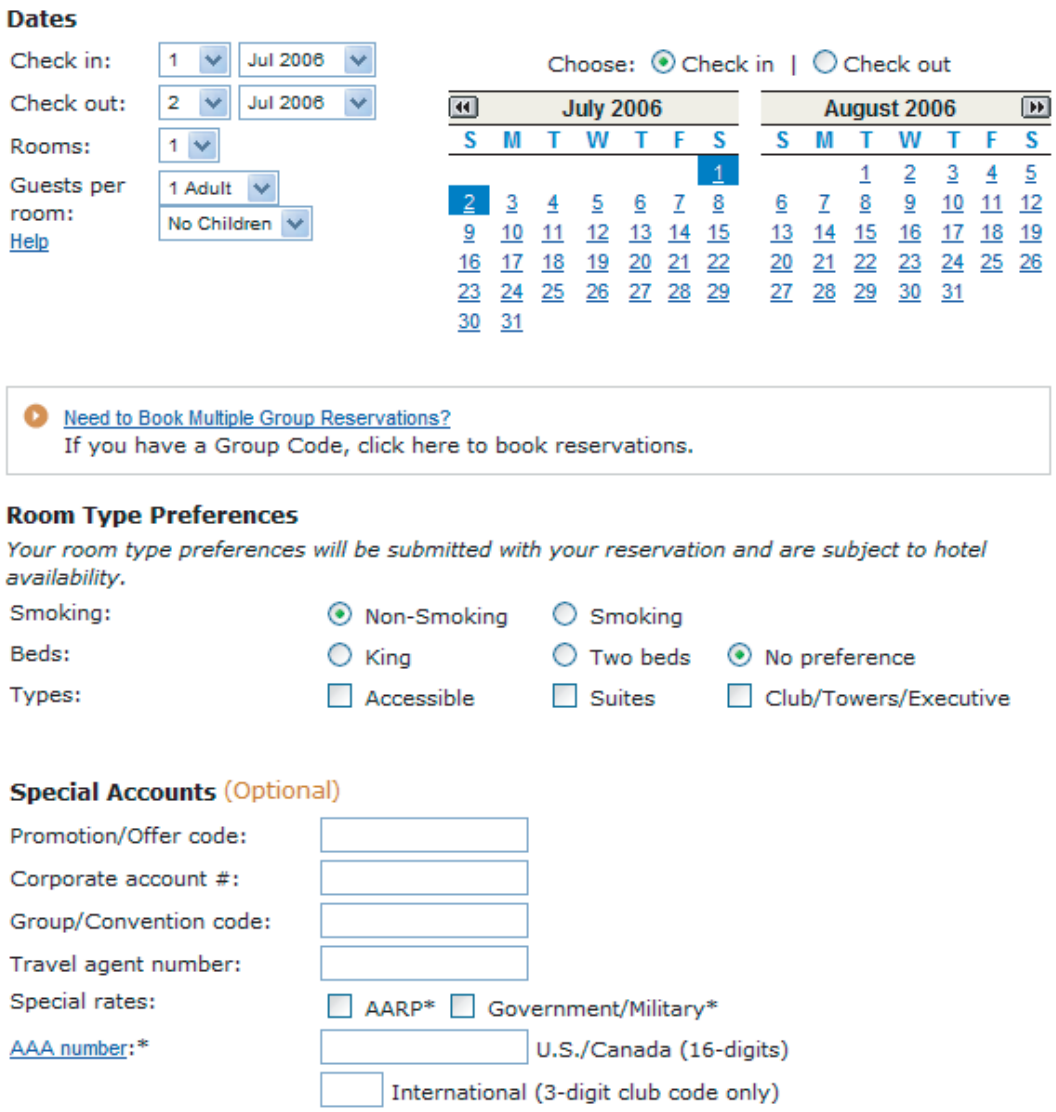


Figure 4.3 **Another form, more acts of finding.** ***Source:*** **Accessed through the online Hilton reservations site.**

well to determine if she has a corporate account number that might apply, and she also wondered whether she should try to get the AAA discount (but she would then need to find her AAA number).

Nancy also did research on several hotel web sites before arriving at the form shown in Figure 4.3. She looked at several possible hotel alternatives, their positions on a map relative to the location of the trade show, their rates, availability of Internet access, workout facilities, and so on. It took a lot of finding. Even providing a credit card number (slot not shown) required a small act of finding to retrieve her corporate credit card from her purse and copy down its number.

Or consider the email exchange I had with someone who was hosting my trip to give a talk at the University of Eindhoven in The Netherlands. In the message depicted (Figure 4.4), reference was made (highlighted) to a hotel and a mobile phone number. The name of the hotel and the phone number were buried in a trail of previous email messages involving not only the host, but also an administrative assistant, in several different threads with different subject headings. The information was all there, somewhere, but the mechanics of getting to it consumed several minutes of time.

You may think of similar examples from your own life. Find a user ID and a password for your account on Amazon or eBay or the New York Times. Assume you have written the information down and know more or less where it is. Even so, you take time to locate this information again. Find a time when you and your friends can meet for lunch. Find an address. Find a phone number—even your own phone number if you have moved and haven't committed it yet to memory. You may have several ways to look things up. But look-up takes time and takes you away from your current task. Moreover, acts of everyday finding—especially efforts to schedule time with other people—can often cascade from one act of finding to several that involve not only the original finder, but several other people as well.

Given the scope and number of acts of everyday finding, we may be less surprised to see estimates from Feldman (2004) that knowledge workers spend 15 to 35 percent of their workday finding information. These estimates may even be low.

Hi William,

as promised some details of your stay in Eindhoven.

When you arrive to the station you can just walk to the hotel (it is at most 5 minutes walk). I attach a map to this message. The hotel is located at the North side of the central square of Eindhoven called Markt, next to the shopping center called Heuvel Gallery (red area on the map). In case you are lost don't hesitate to ask people around, they all speak English and are friendly.

I will book a dinner for us at about 19.00. Let me know if you have any preferences for the food. You can always call my mobile, otherwise let's meet about 18.45 in front of your hotel. Let me know if you also keep your mobile on.

Figure 4.4 **Important information is often buried in an email exchange.**

Everyday situations of finding are far removed from classic situations of information seeking and retrieval. We aren't seeking to learn something new or to fill gaps in our understanding of our world. We're just trying to get on with our daily tasks.

Two characteristics are often true of everyday finding:

- *Finding means communication and coordination.* Nancy needed to provide information concerning her stay at a hotel in Montreal. This meant not only looking at her schedule and the schedule for the trade show she was attending, but also calling her husband for his availability for the weekend after the trade show. Her husband, in turn, looked at his schedule. It is even possible that he might have made his own calls to see whether meetings scheduled might be moved.
- *The information sought is often "small."* Everyday finding often involves look-up of what might be called "property values" for an object or, more simply, statements of fact. Statements are embedded in information items such as a web page, email message, document, or the entry for a person in a contact-management database. But statements are rarely themselves information items to be retrieved on their own.

4.3.1 What can be done to help?

When finding involves communication and coordination among people, at least we now have several alternatives to phone calls or face-to-face meetings. We can email or send text messages. If Nancy needed to reach her husband today, she might even request a short instant messaging session. And her husband might participate in such even as he takes part in a meeting at his work. As part of its review of email, Chapter 10 considers the space of possibilities covered by these modes of communication and others as well. Wikis, for example, as their support matures, have the potential to complement and greatly alter email as we know it today so that problems like those illustrated in Figure 4.4 are less common.

When finding targets smaller pieces of information—statements of fact—there are simple steps we can take now to simplify or even eliminate some acts of look-up (see the "What now for you and me" sidebar for suggestions). Over the long run, we can expect continued improvements in facilities of auto-completion such as we see now to help us with the completion of web addresses, email addresses, and the passwords for some of our accounts. Increasingly sophisticated mechanisms of auto-completion might fill out or suggest easy completions for many of the fields in forms such as those in Figure 4.2 and Figure 4.3. The principle might be "write a fact once, read it/use it many times." Possibilities for a greatly improved, general auto-completion mechanism are discussed in the "What next for tool development" sidebars of both this and the next chapter (Chapter 5). A great deal may be accomplished, for example, through the consistent use, across applications, of structured and semi-structured storage and more granular searches able to flexibly locate statements of fact within such a store.

New modes of communication and coordination; better, more granular search—these and other applications of technology promise to greatly reduce the time spent and frustration felt with acts of everyday finding. But the success of these solutions depends on an understanding

of finding fundamentals. As we see in the next section, any act of finding involves the same basic steps and the possibility to stumble with each step.

4.4 Finding is multistep

There are many techniques of finding and even more names for these techniques. People use browsing, directed search, logical finding, location-based finding, linking, teleporting, and orienteering. Underlying this diversity of techniques is an essential interplay between recall and recognition.

4.4.1 Recall and recognize

Figure 4.5 depicts a dialog between two people. Jerry wants to find out more about a person that he and Kate both met recently. He tells Kate a portion of what he is able to recall, and he then tries to recognize one among the list of names that Kate obligingly provides of people who seem to match Jerry's description. From Jerry's perspective, his attempt to find out more about Tony is a two-step process: first recall; then recognize. Jerry can recall Tony's gender, age, and hair color. By giving this description to Kate, he is able to get back, in return, a list of names of people who match the description, and he is able to recognize the name of the person he wants to know more about.

Lansdale (1988) describes a person's efforts to retrieve information as a similar interplay between recall and recognition. Recall may mean typing in a search string or even an exact address for the desired information. In other cases, recall is less precise and may involve spatial or temporal rather than verbal information. A person may recall in which pile a paper document lies but not its exact position within the pile. Or a person may have a rough idea when an email message was sent or an electronic document last modified. In a second step, then, information items or a representation of these, as delimited by the recall step, are scanned and, with success, the desired item is recognized and retrieved.

The steps of recall and recognition can iterate to progressively narrow the search for the desired information—as happens, for example, when people move through a folder hierarchy to retrieve a desired file or email message or when people navigate through a web site to a desired page. However, as illustrated in the depiction of the berry-picking model of

Jerry: "Who was that person we were talking to last week? You know… at Tom and Mary's party? A man… about 40… dark hair…?"
Kate: "Sam? Dennis? Tony?…"
Jerry: "Tony! That's it… What's his last name? Did he used to work at Tom's office?"

Figure 4.5 **Jerry's conversation with Kate is a give-and-take mixture of recall and recognition.**

information seeking (Figure 4.1), the interplay of recall and recognition is not always a smooth progression through a general space of alternatives to a specific item or item set.

Many different finding techniques have been discussed in the literature, including browsing, directed searching, location-based searching, teleporting, and orienteering. The analysis of finding as an interplay between recall and recognition provides us with one way to compare and contrast these different techniques of finding.

For example, Bates (2002) postulates three generic techniques of finding[9]:

1. In *browsing,* the predominant action is recognition. People browse when they do not have a clear idea of what they are looking for. People also browse when they are unable to recall much (e.g., keywords, content words, properties) that might be used to narrow the scope of the search. Finally, people browse when the space is rich in relevant information so that there is little point in further narrowing the scope of the finding effort. People might browse through a journal or conference proceedings, for example, that directly targets their profession or interest area. In a PSI, the "hot" items of a desktop—items in active use—might be candidates for browsing as a primary means of access.
2. *Linking* occupies an opposite extreme in Bates's categorization. A desired item is fully specified by the information in the link. A full reference to an article is a link. A person might send this information to a librarian, for example, with the expectation of getting back the referenced article and no others. The hyperlink on a web page is also a link, with information that not only fully specifies the referenced web page (via URL), but also an associated mechanism for "jumping" to the referenced web page. People follow links to find information when the space of information is extremely large and the proportion of relevant information extremely low. People follow references to information in a library for example. People follow hyperlinks to information on the Web. People might also be expected to rely more on linking to access the "cold" (archived) information in their PSIs.
3. *Directed searching* can occupy a full range of intermediate positions between the extremes of browsing and linking. The person specifies keywords and gets back a results listing which can then be scanned for the desired item. Since search expressions can be more or less specific, directed searching can range from an activity involving mostly recall to an activity involving mostly recognition. One example is full-text searching of the kind supported on the Web.

Similar to linking or a much targeted use of direct searching, *teleporting* is described in Teevan et al. (2004) as a technique in which people try to jump directly to the information they seek. The emphasis is on recall. Recognition-based scanning is minimal. Such a jump might be accomplished, for example, through the specification of the exact address for a desired information item or through specification of a search string that only the item can match. An opposing *orienteering* technique involves an iterative, stepwise progression toward desired information in which recognition and recall each play an important role.

Orienteering is also used by O'Day and Jeffries (1993) to describe the iterative evolution of librarian-assisted searching. They describe an exploratory process in which intermediate searches are like the steps taken along a winding path through an information space.

[9] Bates uses the word "search" instead.

Consistent with the berry-picking model, the results of previous searches affect the next search request issued: “Like practitioners of the sport of orienteering, our searchers used data from their present situation to determine where to go next.” Recall, then recognition, then recall again. O’Day and Jeffries speculate that the value for the library clients they studied was more in the accumulation of search results than in the final result set.

Related to browsing and orienteering, *location-based finding* is the term for a technique observed by Barreau and Nardi (1995) for the return to information on a personal computer. In a recall step, people take a guess as to location (e.g., the computer desktop or a particular drive or directory or folder) and then scan within that location in an attempt to recognize the desired file. The process is repeated as needed—for example, in order to move through a hierarchy of folders to a desired file.

Note that deciding on a finding technique presupposes a larger context in which other techniques, involving other mixtures of recall and recognition, have already been applied. A recognition-dominant browsing through stacks of books in a library, for example, presupposes an earlier recall-dominant step of getting to the library in the first place. The click on a hyperlink to jump to a web page presupposes an earlier step of recognition in which anchor text of the hyperlink is being used to recognize that the jump is the right one to make.

These labels for finding techniques represent useful characterizations of observed acts of finding rather than precise categories with clear boundaries. An episode of finding may admit to several different labels. More important is that underlying these various techniques is an essential interplay between recall and recognition. Problems in finding, no matter the technique, can often be traced to a problem with recall or recognition.

4.4.2 Problems of recall and recognition and some solutions

Nancy is looking for a travel reimbursement form she has filled out previously to use as a template for completion of her reimbursement request for the Montreal trip. She remembers the business trip she took to Chicago two months earlier and recalls that she created a folder called “trip to Chicago, April.” Naturally enough, she looks in that folder for the file she seeks. But last month Nancy decided that it would be easier to group all travel reimbursement requests into a single folder called “travel reimbursements.” Unfortunately, Nancy has forgotten her own decision during her current finding attempt. This is a failure of recall.

Nancy recovers from this failure and recalls the new “travel reimbursement” folder. However, in scanning through the files of this folder, she does not see the file for the Chicago trip. When she saved the file initially, she was in a hurry and simply saved the file under the default name provided which, perversely, was “non-wages funds disbursement request” (the name of the form on the company web site). The file is there, but Nancy doesn’t see it, and she concludes, incorrectly, that she must have saved the file in a third location. Failure to see the file is a failure of recognition.

As Nancy tries to find the information she needs, what problems does she encounter, and what can be done to help her? Several points can be made.

Help can come through search, tagging support, and auto-completion

First, a desktop search facility might help Nancy (or maybe not). "Travel reimbursement" actually occurs nowhere on the form. And, since Nancy has lots of clients in Chicago, a search on "Chicago" alone will bring back several hundred matching information items.

Second, a supplemental tagging system might let Nancy organize her travel reimbursements forms both ways—as "travel reimbursements" and in association with specific trips like the "trip to Chicago, April." Or better, since Nancy does not have much time to tag, the forms she downloads might already have useful tags like "travel reimbursement." A facility able to search on these tags and on field values such as "Chicago" (as the destination) would make it easy for Nancy to find the desired form.

As part of an ideal in tool support, anything recalled about an information item sought can be used in its retrieval.[10] In some digital memory scenarios,[11] the physical location and the even measures of a person's physiological state are recorded and might be correlated with a person's recollections. A future Nancy, for example, might be able to retrieve the form based on her recollection that she worked on the form on the plane on her way back from Chicago and that she had a headache while she was doing so.

But why is Nancy trying to access an old travel reimbursement form in the first place? What need is she trying to meet? Nancy needs the old form to copy its information concerning arcane matters like "Departmental Code" and "Budget #" to the new form. If there were better ways to give Nancy this information—such as through a more general auto-complete—then maybe Nancy could sidestep the retrieval of the old form and its attendant problems with recall and recognition.

Note though, that the auto-complete mechanism can produce its own problems with recognition and recall. With reference to the form in Figure 4.2, for example, an auto-complete mechanism cannot just list one or more budget numbers as suggested completions—these numbers mostly look alike to Nancy and, even if she thinks she recognizes the right number, she wants to be sure. The auto-complete also needs to include a meaningful "display name" or description for each budget number, much like the display name of an email address or the page title listed along with a suggested web address completion.

On the recall side of auto-completion, what to do about values for fields like "Date/time departed" and "Cost of airfare or train"? Nancy certainly doesn't want a suggested completion for these fields based on past forms! Is Nancy forced to find the information for these fields?

Maybe not. The completion of the travel reimbursement form is part of a larger project. A project, as defined in Chapter 2, is composed of subprojects and basic tasks and may itself be part of a larger project. Projects have goals and constraints and often involve considerable planning. A project, in its planning, execution, and wrap-up, can last for several days to several months (or even years).

[10] See, for example, Lansdale (1988) and Lansdale and Edmonds (1992).

[11] See, for example, Czerwinski et al. (2006) and Gemmell, Bell, and Lueder (2006).

Nancy's trip to Montreal is a project with goals that might include "meet with key clients" and "see what the competition is up to" and "have some fun in Montreal." Components to this project might include "get travel authorization," "make travel arrangements" (a project in its own right), "arrange for meetings with key clients," and so on.

From this project perspective, the values on the travel reimbursement form for fields such as "airfare" and "departure date" are not new at all. Airfare, for example, is specified in the e-ticket that Nancy gets via email from the corporate travel agent. Dates of departure and return are specified on the travel request form (and also in email to the travel agent and also in Nancy's calendar and also in her out-of-office message). Chapter 2 discussed projects as a unit of analysis for the study of PIM. Projects in our practices of PIM can provide a basis for bringing together pieces of information that are otherwise scattered by the various tools (computers, devices, software applications, etc.) we use. We can begin now to organize our information by projects. Tool support can help us do more, more easily, in the future. Projects as a basis for information organization and reuse are discussed in greater detail in Chapter 5.

When information is in several forms, finding can be more difficult

As illustrated in Figure 4.6, Nancy, like the rest of us, may be delayed in finding information, or may experience an occasional outright failure, for the simple reason that she is looking in the wrong place. New tools and applications can make matters worse by increasing the number of possible places to look.

Figure 4.6 **Relevant information may be in any or all of several places.**

In one study reported by Bruce, Jones, and Dumais (2004), participants were asked to return to web pages they had visited up to six months previously by whatever method they chose. If participants remembered having saved a pointer to a target web page (e.g., as a web bookmark or in an email message or a document), they would sometimes first try to find this pointer for use in getting back to the web page. Overall people succeeded in getting back to targeted web pages in 95 percent or more of the trials (depending on conditions) and were generally quick to do so (with times averaging under a minute). But the time-out failures that did occur (after 5 minutes) often pointed to problems of information fragmentation. For example, one participant looked for a web pointer first in her Favorites list, then in selected email folders, then in folders under "My Documents," before finally locating the web pointer inside a presentation she had saved to a network drive.[12]

The information needed to complete a task is often scattered across several different computers, mobile phones, PDAs, and other devices. On a single computer, task- and project-related information may be scattered between different form-specific information organizations—one for email messages, another for e-documents and other files, and yet another for web references. Email messages may themselves be scattered between several email accounts. Documents and other files in the same folder hierarchy may be organized inconsistently by competing organizational schemes. In Nancy's case, for example, older travel reimbursement forms are still organized under folders for the trip to which they apply, whereas newer forms are organized into a single "travel reimbursements" folder.

Recognition failures are also more likely when information is fragmented. For example, a document we can easily pick out in a folder listing may be much more difficult to recognize from its appearance as an email attachment. Even subtle changes in the context of an item's appearance—its appearance in a folder listing vs. in a listing of search results, for example—may hinder us in our efforts to recognize the item.[13]

A multitude of versions and variations can make fragmentation problems worse

The difficulty in returning to information in a fragmented PSI is made worse when only one item will do. We want to respond to the latest message in an email conversation. We need the latest version of a document we are working on. Retrieving and working on an earlier version of this document can mean that the critical changes of the latest document are left behind. The challenge to get back to the right information item in the PSI increases as we keep multiple versions of a document and as still more versions of a document arrive as email attachments from other members of a project.

The specificity of the information need when re-finding items in the PSI stands in stark contrast to the much broader statements of need people often have when approaching the Web.

[12] Finding the pointer in her presentation was mostly luck. She had expected to find it first in the other places visited and expressed confusion and frustration when she did not. As it turned out, a pointer to the web site was in a "Favorite" she had created but in a folder different from the one she looked in.

[13] The dramatic impact that context can have on our ability to recognize is illustrated, for example, in the *encoding specificity studies* of Tulving (1983) and Tulving and Thomson (1973).

Web searches tend to define, implicitly, broad equivalence sets. Any web page in the set will do—that is, the same or similar information is available on several different web pages. Also, especially with no preconceived notions concerning what is or ought to be there, people may overestimate the quality and completeness of results returned by a web search.[14]

A person's challenges of version control increase further if, for some documents, there is no sense of the latest version. Or, rather, there are many, many variations of the document and a desired latest version of each. For example, in my conversations with the CEO of a large financial institution, he described a careful system he had developed in collaboration with his administrative assistant for the repeated use of the same set of PowerPoint presentations with different clients. Each variation was saved as a separate file and given a different name according to the naming scheme <base presentation><client name><presentation date>. Only a few slides changed between variations but these changes were critical. It would not do to have the slides targeting one client in a presentation given to another client!

The increasing challenge to manage multiple versions and variations of a document also points to another, potentially serious, limitation in search facilities. A search may return a long listing of results with each item closely matching the query. The person is then left with the difficult task of deciding which result is the right one for the current need. Finding the right search restrictions to single out the document we want—a recall step—is not at all easy to do; but neither is the recognition step of deciding which document in the results listing is the right one based on differences that are important but not at all apparent.

Fast searches give documents a chance to speak for themselves through their content. Additional terms can be associated with a document and searched as metadata. But we cannot rely on the search of document content to find the latest version of a document, nor should we necessarily trust the accuracy of a tag announcing that the document is the "latest version" (especially not when several such documents are retrieved!).

In some work situations, shared use of a formal document management system may help in version and variation control. However, such systems require extra time and effort to use and impose a level of formality that may not work in many situations of PIM. The CEO above was very knowledgeable concerning software tools and technology and certainly knew of and had used such document management systems. However, his need for management of multiple presentations grew, and became apparent only gradually, as the number of presentations and their variations also grew. By this time, he and his administrative assistant had developed their own system of management through careful use of folder and file names.

If fragmentation is bad, what's the solution?

And . . . is fragmentation necessarily bad? Over the years, several prototypes have explored variations on a theme of information unification: let all information items be of the same basic

[14] See, for example, Blair and Maron (1985) for research suggesting that people tend to overestimate recall rate—the percentage of relevant documents returned by a query.

form—all able to be manipulated in the same basic ways.[15] For example, Haystack represents an effort to provide a unified data environment in which it is possible to group, annotate, and reference or link information at smaller and more meaningful units than the file. In the Haystack data model, a typical document is actually an assembly of many individual information objects. Larger objects can be assembled from small objects as needed, and objects can be dynamically assembled into any number of collections. An email inbox, for example, is just another collection of special "email message" objects.

When such schemes of unification are reviewed in seminars I teach, people often feel uneasy. I liken their reactions to the uneasiness you or I might have if a "this is your life" party were being planned in our honor to be attended by everyone who has ever been a part of our lives—relatives, friends, former lovers, business associates, neighbors—everyone. Do we really want all these people in the same room? People in our lives are naturally separated from each other according to physical location, stage in our life, and our various roles and activities. Separation can be a good thing.

Likewise, with our information, separations by device, application, or email account are sometimes useful. For example, we may keep different email accounts: one for personal email and another for work-related email. Separations can help to divide a vast sea of personal information into manageable regions. Even better is when separations are under our control. We should be able, for example, to remove separations when they begin to hinder rather than help us in our efforts of PIM. And we should be able to assemble task- or project-related collections of information regardless of current separations.

But, too often, our information is partitioned not for our convenience, but for the convenience of tools narrowly focused on a particular function such as note taking or the management of "tunes." We then have trouble grouping documents, photographs, notes, music, and email together according to a larger task or project for which this information is needed. In some cases, we have little reasonable alternative but to use the interface provided by a specific tool as our means of manipulating any of the items created through this tool. This is the case, for example, in the use of Apple iTunes, Microsoft OneNote, or most email applications. Each of these applications defines special forms of information (e.g., songs, notes, email messages). The information items of a form cannot be easily manipulated other than through the application that supports this form.

Other tools, such as applications to create text documents, spreadsheets, or presentations, preserve our direct manipulation of the files (or "documents") as an information form. But even these applications can work in self-centered ways that make small acts of finding more difficult. One sign of this "app-centeredness" is apparent when we want to use the application to save a new information item, open an existing item, or compare two items, and we are thrown away from a folder representing our current task and into an application's default folder as a starting

[15] For a discussion of property-based approaches to unification, see Dourish et al. (1999). For discussion of time in particular, as a basis for unification, see Fertig, Freeman, and Gelernter (1996). For a discussion of a more fundamental unification based on a resource description framework (RDF) representational scheme, see Karger et al. (2005).

point. We want the notes we take and the songs we own to participate in larger activities in our lives. But sometimes we are hindered in our efforts to do so by the very tools that are supposed to help us. At other times, our information is partitioned and scattered by happenstance. Sometimes the chaos in our information reflects a larger chaos in our lives. Regardless, crossing these partitions to gather needed information can be time-consuming and error-prone.

Separations of information that work against us rather than for us and that we can't seem to or don't know how to change are a sign of *information fragmentation.* Information fragmentation is always, by definition, bad.

But decrying the evils of fragmentation does not—by itself anyway—get us any closer to a solution. Better searching, in combination with better, even automated tagging, can be an important tool in the fight against information fragmentation. But search too is limited. A single search can't yet, for example, cross all partitions in our information. And searching to distinguish among multiple versions and variations of a document can be especially problematic. Fragmentation may (or may not) diminish through convergences afforded by developments in our portable devices and on the Web, as explored in Chapters 12 and 13.

Even so, many of the agents of fragmentation are likely to remain a while longer. Whatever our reaction to visions of unification, we're not likely to see these realized anytime soon for reasons explored in greater length in Chapter 14. In the meantime, information will be fragmented in all the usual ways—by location, device, software application, and by us, too, through our own inconsistent ways of organizing our information.

But some of us already have an answer. In seminars, I often informally ask attendees how satisfied they are with their current practice of PIM on a scale from 1 (extremely dissatisfied) to 4 (so-so) to 7 (extremely satisfied). Usually about three-quarters of the people are in the middle, with ratings of 3, 4, or 5. Surprisingly, only about 10 percent of the people usually give a low rating of 1 or 2. That leaves about 15 percent of the people who give a rating of 6 or 7.

What's different about these "PIM satisfied" people? In talking with them, two points are apparent:

1. PIM-satisfied people are already actively engaged in a self-study of their PIM practice. In relation to their tools, organizational schemes, and daily habits—even their choice of friends and colleagues—PIM-satisfied people are constantly noticing what works and what doesn't. PIM-satisfied people actively explore ways of doing things better.
2. PIM-satisfied people are realists rather than purists in their practice of PIM. Not every tool is used or even paid attention to merely because it's new and cool. Not every email message gets answered. Not all information gets organized. PIM-satisfied people pick their battles, so to speak, targeting some information for special attention and letting the rest "flow by."

What is worth keeping and organizing and what isn't? We give special attention to this question in Chapter 5. But first, there's more to say about finding. We have considered recall and recognition as a kind of two-step that is done in various concentrations across all finding techniques. Each is a potential source of problem and failure in finding. Now let's consider two other steps of finding and the problems that can arise with each.

4.4.3 Remember to look

Not all problems with finding can be attributed to failures of recall and recognition. Many opportunities to re-find and reuse information are missed for the simple reason that people forget to look in the first place. In a study by Whittaker and Sidner (1996), for example, participants reported that they forgot to look in "to do" folders containing actionable email messages. Because of mistrust in their ability to remember to look, people then elected to leave actionable email messages in an already overloaded inbox. Inboxes were often further loaded with copies of outgoing email messages that might otherwise be forgotten in a sent mail folder.

Web information is also forgotten. In a study of web use, for example, participants often complained that they encountered web bookmarks, in the course of a "spring cleaning" for example, that would have been very useful for a project whose time had now passed.[16] In the study by Bruce, Jones, and Dumais (2004) mentioned earlier, participants often had bookmarks ("Favorites") pointing directly to targeted web pages to which they were asked to return. Yet these bookmarks were used less than half of the time. Marshall and Bly (2005) report a similar failure to look for paper information (newspaper clippings).

If the old adage "out of sight, out of mind" is true, then maybe the converse is true too: Keep items in view to keep them in mind. Reminding is an important function, for example, of paper piles in an office.[17] Email messages in an inbox provide a similar function, at least until the messages scroll out of view.[18] Barreau and Nardi (1995) observed that users often place a file on the computer desktop in order to be reminded of its existence and of associated tasks to be completed.

Visibility helps. But a person must still be prepared to look. Piles on a physical desktop can, over time, recede into a background that receives scant attention. Likewise, as online advertisers surely know, people can learn to ignore portions of a computer's display. Also, the ability to manage items and keep track of items in view—whether on a computer screen or on the surfaces of a physical office—degrades, sometimes precipitously, as the number of items increases.[19]

Attempts to compensate for the limitations of visible reminders can introduce other problems. People who adopt a strategy of repeatedly checking their email inboxes in order to respond to messages before these scroll out of view (and out of mind) may end up "living" in their email application with little time or attention left to accomplish work requiring sustained levels of concentration. People who immediately click through to interesting web pages, for fear of forgetting to look at these later (even if they bookmark them), may have their session of web use degenerate into an incoherent sequence of page views scattered across a wide range of topics with little to show for the experience.

[16] See, for example, Jones et al. (2002).

[17] Malone (1983) was perhaps the first to note this reminding function of "messy" piles.

[18] See, for example, Whittaker and Sidner (1996).

[19] See, for example, Jones and Dumais (1986).

There are many new ways in which a computer-based device might remind people of potentially useful information,[20] including, for example, the spontaneous execution of searches that factor in words and other elements of the current context (Cutrell, Dumais, & Teevan, 2006). However, these reminding devices must walk a fine line to avoid being either extremely annoying or ignored. These devices, like visible space, compete for a very precious and fixed resource: a person's attention.

4.4.4 Repeat?

In many instances, the need is not for a single information item but rather for a set of items whose members may be scattered in different forms within different organizations. Can you take a job candidate out to dinner next Wednesday? You want to say yes, but answering correctly may depend on finding several widely scattered information items (see, again, Figure 4.6).

1. What does your digital calendar say?
2. Is there anything for that evening on the paper calendar that you share with your spouse at home?
3. Is the trade show you were planning to attend next week? Better check the trade show web site.
4. You vaguely recall making another commitment for the same evening to play bridge with friends. Better check through the email messages on your personal email account.
5. What about your son's soccer match? You really don't want to miss another game. Better check the paper flyer with its schedule of games.

In finding situations such as this one, retrieval of "four out of five" is not good enough. We've all been in situations where we said yes to an engagement only to discover, belatedly, a conflict and then to be in an awkward position of having to cancel one or the other commitment. We thought we checked all possible commitments. We missed only one. Or what about the meeting we arranged or the party we hosted? We thought we invited everyone. But the person we missed is now not speaking to us. Or what about the tax return we filed? In some finding situations, "most" or "almost all" is not good enough.

When all items in a set need to be retrieved, chances of failure increase with the size of the set. Suppose the likelihood of finding any given item in a set of needed items is 95 percent. And let's assume that the likelihood of retrieving each item is independent of the others—an item has a 95 percent chance of being found no matter how many other items have already been retrieved. With these assumptions, the likelihood of successfully retrieving five items (e.g., the five items needed to answer the question above) goes down to 77 percent—that is, the chances of failure are now 23 percent, not 5 percent.[21]

But the chances of success can be worse than expected from the assumption that items are retrieved independently from one another. In situations of *output interference,* items retrieved

[20] See, for example, Herrmann et al. (1999).

[21] The chance of retrieving all five items is 0.95 * 0.95 * 0.95 * 0.95 * 0.95 = 0.77. The chance that at least one item will not be retrieved is 1 – 0.77, or 23 percent.

first may interfere with the retrieval of later items in a set—perhaps because the act of retrieval itself strengthens the items first recalled at the expense of unrecalled items (Rundus, 1971). Some of us may experience this effect when we try to think of everyone in a group of eight or nine friends. No matter whom we list first the last one or two people are often the hardest to remember.

The chances of successfully retrieving all members of a set can also be much better than is predicted by assuming a strict independence of individual retrievals. Every now and then we may find that the exact information needed for the current task comes to us packaged already or can be readily assembled. For this to happen, some forethought is required (by us or others). Support for larger task- or activity-based units of retrieval is discussed further in Chapter 5's discussion of keeping and organizing. But, in these instances, it makes more sense to think of the folder, pile, or collection of items sharing a tag as the unit of retrieval.

Retrieval may also be better than predicted by strict independence if the items we need to retrieve have an organization or are related to one another so that the retrieval of one item actually facilitates the retrieval of other items.[22] One everyday example of what we might call *output facilitation* seems to occur, for example, when remembering the characters of a well-told story or a good movie. The fabric of the storyline helps to connect the characters.

Does your information tell a story? How do the pieces come together? How do they interweave?

And what can our tools do to help? The computers we use have a wealth of raw information concerning our access patterns. How predictable are you in your daily patterns of information interaction? Do you usually open the same applications when you restart the computer? Do you usually go to the same collection of web sites at the beginning of the day to check news, weather, sports, and the stock market? What about the sources you check when looking up a word or a technical term? Do these include reference sources such as Wikipedia, various online dictionaries, or a trusted collection of online magazines, bulletin boards, and blogs?

Trusted tools might begin to construct *information assemblages* based on our patterns of access and use. The applications we routinely open can be preloaded on start-up. The web sites that we routinely consult might be merged into various "mash-ups" (see Chapter 13)—one for start of day, one for look-up, one for when we need an emotional lift.

A tool might even use our access patterns to help us avoid double-booking ourselves. Sources of conflict may be registered in the email messages with date information that we've received and responded to recently, and in the one or more digital calendars we use or reference (e.g., a corporate or organizational calendar) and in the web sites we routinely visit—especially those with dates prominently featured (e.g., conference or trade show web sites).

In short, there are many possibilities, still mostly unrealized, for tools to help us in the "re-collection" of the information we need.

[22] See, for example, Bower et al. (1969) and Jones and Anderson (1987).

4.5 The limitations in ideal dialogs of finding

Wouldn't it be nice if the computer worked like a well-trained human assistant?

In the dialog depicted in Figure 4.5, Jerry is able quickly to establish Tony as the object of conversation. Once this is done, he can ask Kate for more information about Tony, such as his last name or whether he previously worked in Tom's office. Obviously, it helps that Jerry and Kate know each other and speak the same language. It also helps that they have overlapping experiences of events like Mary and Tom's party.

Jerry is smart enough to not tell Kate everything he recalls concerning his encounter with Tony. He doesn't, for example, tell Kate that Tony reminded him, in some strange way, of his best friend in grade school. This would have meant nothing to Kate.

Kate, in turn, tries to be helpful in her responses. She does not, for example, say something like "Do you mean the person who had on a dark-blue sports jacket?" She knows that Jerry is generally oblivious when it comes to details concerning what people are wearing. Kate also doesn't do something silly like speak, in alphabetical order, the names of all men from the party who strictly match Jerry's description. She remembers which people she saw Jerry talking with, for example, and uses this as a basis for ordering her responses. She speaks her best guesses first.

As the shared context between people increases, dialogs can often become extremely efficient, even as these become nearly impossible for a third person to follow. Such dialogs can occur, for example, between two married people, people who have worked together for a long time or even people who are sharing a car on a cross-country road trip. It might even seem, at times, as if we're speaking to ourselves in altered form.

In our interactions with our computers, we're getting closer to an ideal mentioned earlier in which anything recalled about an information item or the circumstances surrounding encounters with it (e.g., time of last use or nearby "landmark" events) can be used to help find this item again. Moreover, our computer as assistant can order its responses so that the most likely object of our interest is always near the top.

We might call this the "me-speak ideal" (or, some might say, the "me-speak extreme") in which talking to the computer (possibly actual talking supported by voice recognition technology) seems like talking to ourselves—only a version of ourselves with a more exact memory for past events. The computer truly could be said to operate as an extension to our own memories (see Jones, 1986).

Here we consider more practical limitations of me-speak.

One practical problem: Not everything we need from our information can be readily articulated in a question or a search expression. The question/response of "Why didn't you tell me?!!?" . . . "Because you never asked me . . ." is a common device for sitcom laughs. But the exchange aptly expresses an underlying truth: we often have difficulty expressing what it is that we want or need.

Of course, a really smart computer—with an understanding not only of us but of the situations in which we need information—might even provide information we need but don't think to ask for. And the computer might do so with the deft, discreet diplomacy of an experienced human assistant—picking just the right times (Horvitz & Apacible, 2003) and just the right mode of delivery ("will you be going to the health club before or after your appointment with . . .?").

But, such intelligence is many years away from reality.

A small step toward me-speak might be a usage-based ordering of the items to be considered during the recognition phases of finding. Put items that are used more frequently at the top of the list; let less frequently used items fall to the bottom of the list. Certainly people do something similar when talking to each other. If Jerry asks Kate again later, "Who was that person I was talking to at Tom and Mary's party?" we would expect that Tony would now be Kate's first guess this time, not her third.

But Teevan, Capra, and Pérez-Quiñones (2007) note that dynamically rearranging a list of item referents in the human–computer interaction based on past experience does not always help and sometimes hurts. To speed up menu access, for example, Mitchell and Shneiderman (1989) tried rearranging menu items so that references to more commonly accessed items would "bubble" to the top of the list over time. However, the opposite happened. Menu access slowed. Because commonly selected items no longer appeared in the expected position in the menu, users were forced to scan the entire list. Similarly, White, Ruthven, and Jose (2002) gave people lists of relevant sentences which were dynamically reranked as the search process proceeded. Users did not enjoy the experience, nor did they perform as well as when the listing of sentences was static. Teevan (2006) completed an analysis of Yahoo query logs that focused on queries repeated by a user. Users were less likely to click through to a web page they clicked to before if its position in the results listing changed. And when they did click, they took more time.

Now what's going on here? Certainly we can think of many instances where we're perfectly happy to have the mostly likely completion listed first. Think of our email application's suggestion of an email address in response to our partial typing of a recipient's name. Or think of a web browser's suggested completion of a web address. We're happy to see the item we want listed first so that we don't have to scroll down the list. Even better if it's highlighted, so we can simply hit return to accept.

One key determinant is the influence of context on our ability to recognize a desired representation for an item based on its representation (e.g., in a list of alternatives). Sometimes context matters; sometimes it doesn't.[23] We know this from personal experience. Sometimes we know exactly what we're looking for. As long as an item's display is reasonable, we'll recognize it no matter where it is on a page or in a listing—better in these cases to list the item where we'll see it first. But at other times, context matters a great deal. In an extreme, our ability to recognize an item is totally dependent on its occurrence in a context of occurrence. We may remember, for example, that the item is "the third one down" or "in the lower left corner."

[23] See, for example, Murnane, Phelps, and Malmberg (1999) for an analysis of when context matters and when it does not.

More often, recognition is driven partly by the appearance of the item itself and partly by our memory for the context of its occurrence. If we repeatedly select an item from one position in a listing—the "Save as" command in a menu listing, for example—we may experience at least a moment's disorientation if this item appears in another position. It's not just "Save As" anymore but "Save As as this occurs in the fifth position." Some of us may have a comparable experience in recognizing other people. It may take us a few extra seconds, for example, before we realize the person we're talking to in the grocery store is actually someone we know from work.

In interactions with our information, space often does matter. A sense of location and context matters.[24] Visibility matters. The visible fabric of information can operate as a powerful extension to our internal, overtaxed, and limited working memories. We're reminded of things we might otherwise forget. We see relationships in our information that we might otherwise miss.

4.6 Wayfinding through the PSI

In their studies, Barreau and Nardi contrasted *location-based finding* with what they termed *logical finding*—roughly equivalent to directed searching as described in this chapter. People overwhelmingly preferred location-based finding as a method of returning to files on their personal computers. Similar to browsing as described by Bates, location-based finding places greater emphasis on recognition but as directed to information on a person's computer rather to a public store. People take a guess as to location (e.g., the computer desktop or a particular drive or directory) and then scan within that location in an attempt to recognize the desired file.

More recently, participants in the Teevan et al. (2004) study—all very technically savvy MIT computer science graduate students—expressed a strong preference for orienteering as a means of finding information, whether local to their computer or on the Web. Like location-based searching, orienteering is shifted toward a reliance on recognition rather than recall. Orienteering can include directed searches, but these are situated and represent small steps in the finding effort rather than large leaps. An orienteering style of return to a web page, for example, might follow the jump to a web site with a site-specific search for the desired web page.

Teevan notes that participants preferred orienteering even when they knew exactly what they were looking for in advance. But perhaps they preferred orienteering precisely *because* they knew what they wanted. Teevan speculates, for example, that orienteering gives people a greater sense of location and context. With a sense of location and context, people can better control the direction of their finding efforts, and they can be more confident that they've reached the right information.

4.6.1 The role of desktop search facilities

The reluctance to use directed search, even in small steps, as a primary means of returning to information in the PSI persists even as desktop search facilities show dramatic improvements

[24] See Teevan et al. (2004).

in speed, ease of use, and integrative scope of search results. In my own ongoing informal survey of people who have installed and use an integrative desktop search facility (e.g., one able to search quickly across files, email messages, recently visited web sites, etc.), people still express a preference for orienteering as a means of returning to information within their PSIs. In results so far,[25] more than 90 percent of the respondents indicate that they used a search facility only as a secondary means of access after primary methods of return such as scanning the inbox or the desktop had failed. Consistent with this informal evidence are results from a formal evaluation of the Stuff I've Seen (SIS) desktop search prototype involving 234 people over a 6-week period (Dumais et al., 2003). Logs of usage indicate that the majority of items accessed through SIS (54%) were last opened over a month ago.

Are you an enthusiast of desktop search facilities? How often do you use your favorite facility? During a seminar I gave in 2005, one participant said that he planned to use a search facility instead of folders as his primary means of getting back to information. When we exchanged email about a year and a half later, he was still an enthusiast and said that he used desktop search as his preferred way to get back to information—"approx. 5 times a day." Five times is a lot—especially if those instances make it possible to access information that he would otherwise not be able to access or would need to spend significant time to access. Undeniably, a desktop search facility can be a powerful addition to the tool set we use to re-find information.

But, for most of us reading this book, five uses of a search facility do not begin to account for the number of acts of re-finding (let alone the total number of acts of finding) in a typical day. Try keeping your own log and count the number of times in a typical day that you:

1. Look for a document either to open and check some fact, reuse some information, edit, or send along to someone else.
2. Look back through your inbox for email messages you have not yet processed, or search for replies to a message you sent out, or look for messages from a particular person.
3. Check and recheck your calendar—perhaps to see what's next or coming up on your schedule, or to search times you might be free for an engagement, appointment, or meeting.
4. Return to various favorite web sites to check news of the day or to look up information on a topic.
5. Look for a song to listen to or a photograph to look at or funny story to read (one more time).
6. Be sure, also, to count incidents of finding like those described above as prompted by the need to fill out a form.

Many acts of everyday finding are so routine as to escape our attention. Yet the completion of these acts of finding takes time and can take our attention away from the work we're trying to do. Email alone can be a significant distraction. As we send one email message or check for one response, how tempting it is to read other email messages as well! Similarly with the Web, in our effort to find or re-find information for our current task, we are often tempted to look at other information as well, even though it has little or nothing to do with the task at hand.

[25] I've asked the question to over 300 people so far in various seminars, presentations, tutorials, and courses.

Getting off track is not always bad and often we may find information that may be useful sometime later. But a problem in our use of an email application or a web browser as a separate standalone application is that we go to the application. In doing so, we leave the context of our work (even if the document prompting our excursions remains in view), and we enter the world of the web browser or email application. These problems diminish as finding activities directed toward email, the Web, and other information are increasingly integrated with one another and into the working situations from which information needs arise. But we have a long way to go.

4.6.2 Wayfinding as a round-trip

The reasons for a reliance on browsing, location-based finding, and orienteering are aptly subsumed under another term: *wayfinding*. Peter Morville, in his book *Ambient Findability* (2005), defines wayfinding as "knowing where you are . . . knowing your destination, following the best route to your destination, being able to recognize your destination and *finding your way back to your starting point*" (p. 17, emphasis added; see also the "Wayfinding" sidebar by Peter Morville in this chapter).

Another characteristic of wayfinding is the ability to reproduce the experience again later. The wayfinders who located a new uninhabited island wanted not only to return to their home island, but also to guide their fellow islanders to the new land. A person who gets to a specific web page, email message, or file on one day might like to be able to repeat the experience the next day.

The term "wayfinding" encourages a subtle shift of emphasis in finding. Certainly the destination—getting the needed information—is important. But also important is a successful return, with information "in hand," to the context where the information is needed. Also important is the experience of the journey. Is finding easily done in the context of our work, or does finding take us away from our work and down paths having little to do with the task at hand so that we leave our work at the end of the day frustrated for the lack of progress? Does finding feel like a matter of providence over which we have little control, or do we have confidence that we can reliably, repeatedly get back to information as needed?

The term "wayfinding" was coined by Kevin Lynch and used in his book *The Image of the City* (1960). Lynch developed a rich vocabulary to describe the features of a city that impact people's ability to wayfind:

- *Paths* such as streets and walkways connect the parts of a city together.
- *Districts* are major sections of the city with common identifying character—think, for example, of the SoHo and East Village sections of Manhattan.
- *Edges* such as walls, fences, and other barriers help define and separate, but also relate, different districts.
- *Nodes* such as street corners, squares, and subway stations serve as points of reference and transition.
- *Landmarks* such as large churches, skyscrapers, and museums serve as important points of reference.

PSI counterparts for these elements of a city either don't exist or they exist imperfectly:

- The paths that make it possible to move between documents or between web pages are represented implicitly in constructs such as a folder hierarchy or a hyperlink. But paths are generally not handled as objects in their own right, nor are they given meaningful, visible features. For example, a hyperlink in a web page changes color to indicate that it has recently been clicked. But people are given no indication of how frequently the hyperlink is clicked. People cannot see how the paths they take—to web pages, documents, application commands, and so on—are routinely, repeatedly combined over a period of time. Virtual paths do not deepen or widen with use.
- The concept of districts is partly represented in the folders we keep and perhaps also in various clusters of desktop icons. Folders as districts are not well integrated. For example, email folders contain only messages, and file folders generally contain only e-documents. There is little to express the size, character, or composition of a folder other than its name or possibly color.
- Folders generally have hard and fast boundaries—an item is either in a folder or not (although boundaries are softened somewhat by recent efforts to create "search" or "smart" folders). But these boundaries are not really edges that can serve to relate as well as separate. Folders are related to one another primarily by a parent–child relation. Even something as basic as ordering, as a way to relate sibling folders, is poorly supported.
- A folder in its relation to its subordinate folders can be seen to act as a node. Click on the folder to see and choose among its subfolders. Likewise, a major web site, as a hub or starting point/meeting point, acts as a kind of node. Are there others?
- What counts as a landmark in the digital regions of a person's PSI? The Start button? Desktop shortcuts? Applications with icons always in view? The home page icon and other icons always in view in a web browser?

As summarized in Table 4.2, wayfinding constructs identified by Lynch work well for the Manhattan borough of New York City but seem to stretch or break as we try to apply them to a PSI. Do gaps in the mapping of wayfinding constructs to a PSI point to basic limitations in the metaphor, or do these gaps represent opportunities to improve PIM?

Dillon and Vaughan (1997) review literature suggesting that, in navigating physical spaces, people form three distinct types of mental representations over time: landmark, route, and survey. Representations are generally seen to build on one another. For example, we might navigate first in New York City according to major landmarks like the Empire State Building or the Chrysler building in order to get to a friend's house from the 34th Street PATH station. Over time we develop a route that we follow to get to our friend's apartment. Survey representations of the kind needed to form a mental map of the environs emerge later. As we acquire a survey representation, we can begin to explore alternate routes to our friend's apartment that are shorter or that take us, for example, past a corner store where we can buy last-minute food items for dinner. Alternatively, of course, we can get a map of Manhattan.

Table 4.2 **Wayfinding constructs make sense for Manhattan, but do they make sense for your PSI?**

	In Manhattan . . .	In your PSI?
Paths	Park Avenue, 5th Avenue, 34th Street, 45th Street, the Brooklyn Bridge and its pedestrian walkway, Broadway	Well-trod "paths" through your favorite web pages? The series of clicks to access a network?
Districts	SoHo, Tribeca, Chelsea, the Upper East Side, the Lower East Side, Chinatown, Little Italy	Clusters of icons on a computer desktop? Folders? PICs?
Edges	Canal Street, Central Park, Avenue of the Americas	The boundaries between folders and disk drives? These are separate "districts" of information; but how are they related?
Nodes	Times Square, Washington Square Park, Grand Central Station, Pennsylvania Station	Folders? Favorite web sites? A task bar?
Landmarks	The Empire State Building, the Chrysler Building, the Brooklyn Bridge	The home page icon of a web browser? The "start" button?

However, Dillon and Vaughan question the utility of the landmark, route, survey (map), and other spatial concepts of our three-dimensional physical environment when talking about information. They note, for example, that "what constitutes landmarks in information space remains unclear" (p. 96). Perhaps the information in a PSI and its uses are too fluid, too dynamic, too multifaceted to be adequately characterized through the application of static, stationary concepts such as paths, edges, districts, nodes, landmarks, routes, and so on.

And what about the "way" in wayfinding? Sure, destination is important. And so too is a "safe" return to the situation that prompted us to get information in the first place. But how important is the way to and back?

For that matter, how important is the way to and back as we move through physical space? Most of us probably travel by plane rather than train, automobile, or ship in order to reach a remote destination. We do so for the savings in time even if the journey is often unpleasant with long lines at security checkpoints, cramped seating, and terrible food. At least the journey is short. If we could safely teleport to save even more time, we probably would. The uncertain promise of serendipitous encounters on a longer journey is not likely to outweigh the sure thing we have in the savings of time.

Wayfinding

– Peter Morville

"Wayfinding" as a discrete term originated within the context of what architects call the built environment. First used by the architect Kevin Lynch in 1960 to describe the role of street numbers, directional signs, and other "way-finding" devices in cities, the word was later appropriated by biologists, anthropologists, and psychologists to describe the behavior of animals and humans in natural and artificial environments.

Most recently, wayfinding has been applied to the study of user behavior within digital information environments. We talk about people getting lost in cyberspace. We create "breadcrumbs" and "landmarks" to support orientation and navigation in web sites. While these spatial metaphors are often taken too far, there is no doubting their resonance.

We do import our natural wayfinding behaviors and vocabularies into digital environments, and for that reason alone the history of wayfinding is worth our attention. But at the intersection of location awareness and ubiquitous computing, we are increasingly navigating hybrid environments that connect the physical and the digital. The history of wayfinding only grows more interesting with each step into the future.

But before we lavish attention on *Homo sapiens*, it's worth taking a look at the wayfinding skills of a few other species with which we share planet Earth. Their solutions to the challenges of orientation and navigation can illuminate our own. For example, have you ever wondered how ants find a feeding site and then return home? Lacking maps and street signs and cell phones, these tiny creatures regularly travel thousands of times their own body length to arrive at a pinpoint goal.

After decades of research, behavioral biologists have begun to figure out how. Studies show that ants use a combination of geocentric and egocentric techniques. Geocentric navigation relies on external environmental cues such as landmarks and any available map information. Egocentric navigation relies on self-awareness of distance and direction traveled and is independent of the immediate surroundings. Of course, these senses are imperfect, and errors can rapidly accumulate during the course of a trip. It's the sophisticated combination of strategies that allows for error correction and ultimate wayfinding success.

Sight. Hearing. Touch. Smell. Taste. We're often intrigued by the novel application of these five senses. Bats and whales and dolphins use echolocation to "hear" their way through low-visibility environments. Salmon rely on a powerful sense of smell to sniff out routes as they navigate back to the upstream waters where they will breed. We're also impressed by unfamiliar wayfinding senses such as the polarized vision of ants and honeybees or the biomagnetism of sea turtles, lobsters, and newts. We can't help but speculate what it would be like to possess these remarkable capabilities.

Of course, we've developed some pretty sophisticated wayfinding tools ourselves. By enhancing our natural abilities with such technologies as maps and compasses, we've turned the whole planet into what Kevin Lynch might call a legible environment. But everything is relative. Despite the ready availability of maps and street signs, we still manage to get ourselves lost. Lost in cities or inside buildings or on the way.

Fortunately we appear poised on the brink of a breakthrough. After eons of bumbling around the planet, we're about to take navigation to a whole new level. Wayfinding 2.0. And it begins with location awareness. The crown jewel of next-generation wayfinding is the Global Positioning System (GPS), a satellite-based radio-navigation system that enables land, sea, and airborne users to determine their three-dimensional position (latitude, longitude, altitude) and velocity. Equipped with a GPS receiver and map database, we can find our way like never before. Our kids will wonder how we ever survived without GPS, and not just in the car. GPS receivers and mobile devices that rely on other technologies such as Wi-Fi, Bluetooth, Ultra-wideband, and RFID grow smaller and more popular every year. Handheld units are increasingly common. No more printed maps. No more getting lost.

This is the promise of ambient findability, a world at the crossroads of ubiquitous computing and the Internet, in which we can find anyone or anything from anywhere at any time. It's not necessarily a goal, and we'll never quite reach the destination, but we're surely headed in the right direction. And wayfinding will never be the same.

For information too—notwithstanding the Teevan et al. (2004) results—wouldn't we teleport for the savings in time if we knew exactly what information we needed and we could reliably get to this information and back again? In cases of form-filling, we're happy for no journey at all. The rub of this question, of course, is in the "ifs." Needed information as a destination is rarely as certain as, say, a trip to Boston. Recognizing the needed information is often heavily influenced by the context surrounding its access. Moreover, need is itself revised and refined in the course of our efforts to find information. The way does matter.

Also, there is reason to believe that Lynch's wayfinding constructs do have application to a PSI, though perhaps in ways less direct and less literal. For example, in our work on the Keeping Found Things Found (KFTF) project, we see considerable longevity in the top-level organizations people impose on their information. Top-level folders, for example, often reflect projects, areas of interest, roles, and responsibilities that endure for years (Jones, 2004). Do these count as landmarks?

There is certainly evidence that people are creatures of habit in their access to information, taking the same sequence of steps, or the same route, each time they need to access an information item such as a web page or a file.[26] Many of us may have the experience of following the same sequence of clicks to reach a particular web page—the white pages for address and phone number look-up, for example—even though we suspect there are shortcuts. Do these count as paths? Would a web map showing our current location and likely destinations give us the courage to take a shortcut?

Maps of the physical world can have tremendous value for their ability to instill survey knowledge that might otherwise take years for us to acquire. Maps often appeal not only for their

[26] See, for example, Capra and Pérez-Quiñones (2005).

utility but also their aesthetics. A well-designed map can be a thing of beauty to be appreciated in its own right.

What would a map of our PSI, showing our daily activities of PIM, look like? What paths or "sea lanes" do we trace through our PSI? Where do managed personal information collections rise up as islands of relative structure and coherence? And how would we use this meta-information to improve our practice of PIM? These are fascinating questions that don't admit to easy answers.

I find that a great part of the information I have was acquired by looking up something and finding something else on the way.

U.S. journalist Franklin P. Adams (1881–1960)

4.7 Looking back, looking forward

In looking back over this chapter's exploration of finding, re-finding, and the movement from need to information, several points can be noted:

- *Re-finding is different from finding. Re-finding in a personal store is different from re-finding in a public store.* Finding activities vary according to whether we've previously experienced the information we seek and whether the information is under our control or "out there" in a public store such as the Web. Efforts to re-find information are more focused than efforts to find new information. We're looking for a specific information item and not just any information will do. In our efforts to re-find an item, we should be able to use not only our memories for the item itself, but also our memories for our encounters with this item. Failure to re-find an item in a personal store is different from failure to re-find an item in a public store. The failure can stand for a larger failure of control. If we've lost this, what else have we lost?
- *The journey from need to information is round-trip.* Information, once found, must be applied in the situation that generated a need for this information in the first place. How quickly can we return to this situation? How quickly can we reestablish context? Is the information found in the right form to be used? Also, our assessment of need is dynamic and changes with every step of finding we take.
- *Small acts of finding can kill a day.* A day is often composed of many, many small acts of finding. Find the right budget number to put in an expense form. Find the email containing a phone number. Check our calendar—once, twice, many times. Look up our account name and password for an online service. Even if small acts of finding nearly always succeed in getting the needed information, and even if each takes only a few minutes to complete, these can add up to much of a day's time and its frustrations.
- *Finding is a multistep process with a possibility of stumbling at each step.* What we *recall* about the information we need could be wrong, so that we look in the wrong place or search on the wrong words. Or we may fail to *recognize* an item even though it's "right there." We may fail to *repeat* the finding process enough times to assemble the complete set of items

needed for a decision. And we may fail at the outset because we do not *remember* to look.

- *Wayfinding and finding as a dialog* with a trained assistant (or the computer as a "me-speak" extension to ourselves) each provide unique, useful perspectives as we look for ways to provide better tool support for finding. Each perspective also has its limitations. The dialog falls short in the many cases where we don't know exactly what it is we're looking for or when we may not even think to ask for the information in the first place. And the dialog as a model for support of finding may fall short in cases where our ability to recognize a needed information item, and perhaps to be sure we have the right version, depends on a representation not only of the information item but also of the larger context in which the item usually occurs.

The wayfinding metaphor gives emphasis to finding as a kind of journey through the PSI—from a need and the situation prompting this need to information and then back again. Wayfinding may be a means for us to increase our understanding of and control over our PSIs. And, as with any journey, a journey through the PSI can be serendipitous, yielding useful information we did not expect and would not have thought to ask for. But in cases such as form filling, where we know exactly what information we need, we're happy to dispense with the journey. Just the facts, please.

What Now for IT?

– Bob Boiko

Today's top IT talent wrestles daily with the issues of fragmentation, and they are finding no easy way out. On one hand, there are constant calls for greater integration of information sources. "Why," users ask, "can't I search across email, the Web, and my hard drive?" On the other hand, people also call for more user-friendly systems that require little or no training to use. When IT groups create monolithic systems that access all information, users complain that they are too complex, hard to use, and unreliable. When they opt for simple, single-function systems that are easy to use, users complain that information is "trapped" within the systems and not accessible without special effort.

A similar dilemma arises when IT departments attempt to make information more findable by introducing universal storage and categorization standards (such as exist in document management systems). They craft such standards in order to satisfy user demands for more findability. What they get when they attempt to implement the standards are user demands for a simpler system that does not require them to change their well-worn ways and spend time categorizing the information they are storing.

Users can't be blamed for wanting to have their cake and eat it too. IT staff can't be blamed for trying to give users what they seem to want, even if that results in a contradiction. Rather, the up-front cost to the user (in time and training) of making information findable has to be balanced against the value of making information more available. Of course, to balance this equation, IT workers must be able to calculate the value of finding information faster—something they have not yet been able to do.

Also, wayfinding constructs—such as paths, districts, edges, nodes, and landmarks, with ready application to the physical environs of a city—apply at present (or at best?) in only limited and substantially altered ways to a PSI. These problems of application may point to limitations of the metaphor, or they may point to limitations in our current ways of thinking about and designing for our PSIs.

For example, we follow what might be termed "well-trod paths" through our PSIs in our use of various applications and in our use of the Web. But our PSIs, in their supporting tools, retain little impression of our habits; nor do our current tools make much use of this information to suggest better paths or shortcuts. Maps of our PSIs, like maps of the physical world, may have tremendous value for their ability to instill survey (overview) knowledge that might otherwise take years for us to acquire. And these might be an object of beauty too. But it is not clear what such maps should look like or how they should be constructed.

This chapter has given special focus to efforts to re-find—that is, to find information already (if only briefly) experienced. The chapter has further focused on efforts to re-find information inside our PSIs in stores (such as our computer's hard drive or even the filing cabinets of a physical office) that we ostensibly control. We must remember to look in the first place. We must recall enough about the information to give reasonable scope to a follow-on effort to recognize the specific item we seek. We may need to repeat the steps of recognition and recall several times in the search for a specific item. And a larger episode of finding may require that several items are retrieved to form a complete set.

But, the journey to re-find information does not properly begin with a perception of need or a remembrance of a relevant information item. The journey begins with earlier acts to keep and organize the information in the PSI. What we do at this stage often determines whether the journey will ultimately succeed or fail. This is the topic of the next chapter.

What Next for Tool Development?

– Mike Kelly

Much of the tedium in everyday finding comes from the repeated look-up of small pieces of information. Consider:

- **We may look up a budget number once and then, later the same day or the next, we may need to do so again.**
- **We send our cell phone number to people via email so that they can reach us later if necessary—not once but many times for different people or to the same person at different times.**
- **We spend many minutes sifting through email in the inbox to locate a phone number someone else has sent us. When we received it, we had the idea to make an entry**

for this person in the contact database we maintain, but there was no time. Now the email message is pushed far down in the inbox, and we're taking even more time to re-find it.

In situations like these, a little bit of computer smarts can go a long way. Variations on auto-complete or auto-fill already support the completion of web addresses, email addresses, and passwords. Some email applications now do limited parsing of message text to recognize dates (for suggested meetings, for example) and support the option to enter this in a digital calendar.

But current auto-complete facilities are limited in several ways. Support for auto-complete is fragmented, and so it works differently in different places. If just once we accidentally provide the wrong information—the wrong email address, web address, or password—auto-complete facilities are equally happy to propose this incorrect completion, and ways to correct this are non-obvious and inconsistent. Finally, to the extent applications do maintain auto-complete data today, it is almost always per-machine, which does nothing to solve the problem of synchronization across our different devices: work computer, laptop, home computer, and mobile phone.

It is time to consider a single, unified, and smarter auto-complete facility that can be accessed from all our machines and that works consistently across multiple applications. At the core of this would be a database with objects such as "person" and "budget" and associated properties such as "cell phone number" and "current budget amount." Email applications, word processors, web browsers, and other applications could access this either to store new information or to retrieve information.

The use of auto-complete to store (keep) information is discussed at greater length in Chapter 5. The consolidation of storage here also carries maintenance benefits (updates and corrections are much easier), which are discussed further in Chapter 6. Here let's consider how this newer, higher-functioning auto-complete could help people to find and use information:

- We type "my cell phone number is" and the application is ready to complete with the phone number. The same completion can happen whether we're typing a document (e.g., an invitation) or sending an email message.
- We type the first part of a web address and the application completes—similar to today. But it might also recognize a"well-trod path" through a sequence of web pages and so suggest a jump to the eventual destination. If we always go through pages A, B, and C to get to page D, for example, and we're currently viewing page A, auto-complete might at least suggest page D along with page B. This is tricky to get right, and we have to avoid the famous "It looks like you're writing a letter" problem.
- Budget numbers, directions, and addresses could all become pieces of information readily accessed from auto-complete and inserted into a current information item (whether document, form, or email message).

One key to success here is simple application programming interfaces that support the consistent use of this information. It is especially important that different applications storing information in the system auto-complete add to and reuse existing objects wherever possible rather than create new custom ones. Ideally, we would like all information we have about "Robert Smith" to be consolidated in a single entity—containing his cell phone, his web page address, and so on. How these higher-level objects might work, and what they might enable, is a topic for further discussion in Chapter 5.

What Now for You and Me?

Practical suggestions apply to each essential step in finding.

1. ***Reminding/remembering to look.*** **Look around. What do you need to remember to find? Look at your *attentional spaces* (see also Chapter 5)—for instance, your desktops, physical and digital, your inbox(es), your calendar(s). What have you forgotten to look for? Many of us are in the habit of doing this at the beginning of the day. Also, look periodically through at least the top levels of your various organizations of paper documents, digital documents, email messages, bookmarks, and the like.**

2. ***Recall.*** **If you have information in several places—on different computers, different drives or shares, or in different email accounts—then make a list of these that you can easily consult. Do you need to find the correct version or variation of an item? Then hopefully you already have in place at least an informal process of version and variation management. Think broadly about the places you might look. Use your friends and colleagues as information sources too (and reciprocate). For example, rather than sifting through old email messages for a pointer or a phone number, consider simply asking the sender again.**

3. ***Recognition*****. Look once. Look again. It seems such an obvious and unlikely mistake not to recognize the information that is in view. But this happens to us all the time. In the longer range, of course, a failure to recognize may best be fixed in the manner of an item's keeping—for example, with better names for documents or a name of your choice for a web bookmark. The subject line of an email message, as the counterpart to the name of a document or file, is worth a few seconds of crafting before you send it out so that you are more likely to recognize and attend to replies that arrive in the inbox later. In a group collaboration, some minutes spent agreeing on subject-line conventions can be an investment that pays for itself many times over.**

4. ***Repeat?*** **What else? Where else? Be clear on whether a complete set is really needed. There is no "complete set" of information anytime the Web is involved. We may spend too much time chasing an illusory complete set of information needed**

to write an article or purchase a product or plan a vacation, even as we overlook completion with respect to more mundane matters such as scheduling a meeting.

Some other suggestions:

- ***Get a desktop search facility*** **if you don't have one already. Use it. They are free and generally "well behaved" in their uses of your computer's resources. But recognize its limitations. The search facility you acquire has its place as one of several tools supporting several methods of finding.**
- ***Note the bits and pieces of information*** **you find and repeatedly re-find. Write down for easy reference the budget numbers, phone numbers, departmental codes, names, and the like that you repeatedly need. If you are organized enough to have a database of contacts, then use it (for example, make a budget or a department into a "contact"). But for information we repeatedly access but never seem to commit to memory, it may be more effective simply to write the information down on paper in a notebook close at hand or in a "scraps" file that can be read by a software application we nearly always have open such as a word processor or a spreadsheet application.**
- ***"Begin with the end in mind."*** **This maxim from Covey (1989) applies with special force to activities of finding. Where will you need the information you're trying to find? In what form? Stepping back further, what need are you trying to meet? Not all needs are equally important, and some apparent needs take care of themselves in good time without your intervention.**
- ***Become a student of your finding activities.*** **When finding fails, ask yourself, How could I have done things differently? The answer may lie in better execution of one of the finding steps listed above. But many of the problems of finding information originate and are best dealt with during the keeping and organizing of the information. The "What now for you and me" sidebar for the next chapter deals with that.**