A Review of Organizational Structures of Personal Information Management

Indratmo and Julita Vassileva
Department of Computer Science, University of Saskatchewan
Saskatoon, SK S7N 5C9, Canada
Email: {j.indratmo, julita.vassileva}@usask.ca

Abstract

Personal information management (PIM) covers a large area of research fragmented into separate sub-areas such as file management, web bookmark organization, and email management. Consequently, it is hard to obtain a unified view of the various approaches to PIM developed in these different sub-areas. In this article, we synthesize and classify existing research on PIM based on the approach used to organize information items. We classify the organizational structures into five categories: hierarchical, flat, linear, spatial, and network. We discuss the strengths and weaknesses of each structure along with examples showing how to deal with the weaknesses. Finally, we provide design recommendations and a framework for researchers to experiment with various ideas for developing novel PIM tools.

1. Introduction

Personal information management (PIM) refers to users' activities in acquiring, organizing, retrieving, and processing information in their personal information spaces (Teevan et al., 2006). As part of their daily activities, users create new documents, receive and send email messages, manage appointments and to-do lists, and retrieve information from personal collections and other resources. With the declining prices of mass storage devices, users can store a lot of information items in their collections, eventually exceeding their capacity to manage the items effectively. As a result, they often have difficulties in organizing their collections, in finding needed information, and in using information to achieve their objectives (Bellotti et al., 2005; Malone, 1983; Ravasio et al., 2004). Such difficulties decrease productivity, as users have to spend a lot of time managing information items instead of processing and using the information to accomplish their tasks. Since PIM is integral to the everyday lives of many people, improvement in the design of PIM tools will have significant impact on human-computer interaction.

Personal information in this context does not necessarily refer to information about users, such as their names, addresses, marital status, and occupations. Instead, it refers to information owned or managed by individual users, for example, spreadsheets, email messages, contact lists, calendar entries, to-do lists, and web bookmarks. Personal information, however, is not limited to digital items only, but also includes tangible items such as books and magazines. In this article, we will refer to such personal information as information items or documents interchangeably.

PIM research usually focuses on a specific subject, such as email management, web bookmark management, or file management. Since the research is fragmented, it is hard to see the underlying principles of the existing approaches to PIM. In response to this problem, we provide a unified view of approaches to PIM based on their organizational structures. There are two main contributions of this review. First, we discuss the strengths and weaknesses of each organizational structure (section 2) and their implications for the design of PIM tools (section 3). Second, we identify various aspects of PIM research to give directions for future work (section 4). Throughout our discussion, we provide some pointers to technologies that have potential to improve PIM tools.

2. Approaches to Personal Information Management

In this section, we review existing approaches to PIM and classify them based on their organizational structures. We classify the approaches into five categories—hierarchical, flat, linear, spatial, and network—and discuss the strengths and weaknesses of each structure.

2.1. Hierarchical

Hierarchical approaches use a tree structure to arrange information items. A node in a tree represents an information item (e.g., a file, an email message) or a collection of items (e.g., a file folder). A hierarchical structure is currently the predominant way to organize information items. Many software systems such as file systems and email clients use this structure and allow users to create folders and subfolders to facilitate information classification, management, and retrieval.

One of the strengths of a hierarchical structure is that people are familiar with it. From universities to companies to government, people can find a hierarchical structure used to facilitate the workflow of the organizations. People deal with a hierarchical classification scheme while looking for books in libraries. Most operating systems use hierarchical file systems to manage and organize files. Therefore, organizing information items hierarchically is an intuitive, familiar process.

A well thought out hierarchy can help users understand an information collection better. In a hierarchy, information items relating to one another can be logically structured, and their relationships are explicitly captured. For example, when users have subfolders called “budget” and “team members” within another folder “web project,” they can infer that “budget” and “team members” are part of the “web project.” Even if they have other subfolders “budget” and “team members” in another branch of the hierarchy (part of another project), they will have no difficulty in differentiating between subfolders that belong to the “web project” or the other project, because the relationships between these folders are clear. In practice, people make use of such explicit relationships in a hierarchy to maintain overviews of their collections: whenever necessary, they create a new subfolder and use the folder to keep an overview of related items within it (Ravasio et al., 2004).

People tend to organize information items based on the task relating to the items (Barreau, 1995; Kwasnik, 1989, 1991): for example, grouping all materials for teaching a programming course. The importance of the corresponding task to information organization is reflected in the user practice of using hierarchies to facilitate task management. Specifically, people create folders and subfolders relating to the tasks and subtasks that must be done in a project (Jones et al., 2005). The initial hierarchy of a project might be simple. However, after the user has gained better understanding of the project, the hierarchy becomes more elaborate. In this context, hierarchies serve as a medium for task management—an integral part of PIM.

The practice of using hierarchies as an aid to task management leads to the idea of structure reuse (Boardman et al., 2003; Jones et al., 2005). They propose to enable users to copy the structure of a hierarchy. This function is particularly useful when users have similar tasks or projects repeatedly. For example, the nature of teaching university courses is more or less the same: one needs to develop course notes, assignments, quizzes, a midterm exam, and a final exam. For good organization, a lecturer may put these course materials into different subfolders, resulting in a hierarchy. If users can reuse the structure of a hierarchy, then they do not have to create the same structure manually when assigned a similar task. Furthermore, a structure can remind users of the necessary tasks to do. For example, a subfolder "Quizzes" may remind a lecturer to incorporate a quiz component in a syllabus.

A hierarchical structure facilitates information search and retrieval by enabling users to reduce a search space and to eliminate any ambiguity of a term which can refer to different contexts. Consider the hierarchy of subject categories used to classify books in libraries. While looking for a Java programming book in an online library, users can navigate through the computer science branch and explore the subcategories within it. Within this branch, they do not have to deal with non-computer-science books, which are irrelevant to their information needs. Books about the island of Java (part of Indonesia), for instance, will not be found in the computer science category, because those books belong to a different branch of the hierarchy (geography).

Despite its advantages, a hierarchical structure has some drawbacks. Creating a hierarchy and classifying information accordingly is a heavyweight cognitive activity (Lansdale, 1988; Malone, 1983; Ravasio et al., 2004; Whittaker and Sidner, 1996). This difficulty especially applies to knowledge workers, such as analysts and researchers, who have unstructured and dynamic tasks. Knowledge workers have to deal with ever-changing work contexts and sometimes consider filing documents unimportant (Kidd, 1994). When filing an information item, they may put the item into a category that is appropriate at that time. However, when their goals and work contexts change over time, the hierarchies that they created may become outdated and require a lot of time and efforts to maintain. Furthermore, since they handle multiple tasks concurrently, they may encounter an information item that is relevant to several tasks, making it even more difficult to choose in which category they should file the item. Although people can alleviate this problem by creating multiple copies of a document (or shortcuts), they tend to put a document into a single category only (Mander et al., 1992).

While working on a project, people use and collect information from various resources. They may communicate with their colleagues using email, create a budget for the project on a spreadsheet, and keep relevant web resources in their browsers' bookmarking facility. In this way, information items relating to the same project are scattered in various PIM tools: email clients, file systems, and web browsers. These tools usually maintain their own hierarchies and do not facilitate users to group heterogeneous but relevant items together—a problem known as information fragmentation (Karger and Jones, 2006; Ravasio et al., 2004; Whittaker et al., 2006). Information fragmentation leads to repeated efforts to organize different types of information items (Bergman et al., 2006; Boardman et al., 2003): these studies find that users create similar hierarchies across PIM tools, particularly across email clients and file systems. Consequently, users have to consult various storage locations or PIM tools to retrieve needed information, and it is hard to maintain consistency in overlapping hierarchies.

In response to the limitations of a hierarchical structure, some researchers have extended the notion of a hierarchical structure as follows (Dourish et al., 1999; Karger and Quan, 2004). First, an information item can belong to multiple collections (i.e., categories or folders). The content of a collection can be determined by both static and dynamic components of the collection. The concept of fluid collections (Dourish et al., 1999) and semantic directories (Gopal and Manber, 1999) consists of three components: an inclusion list, an exclusion list, and a query. An inclusion list contains a list of items that must be included in a collection regardless of the query component. An exclusion list contains a list of items that must be excluded from a collection even if they match the query. The last component, a query specifies the criteria of items to be included in a collection. By manipulating these components, users can create various types of collections. They can define simple, static collections that work just like ordinary folders by specifying the inclusion list only. Whenever necessary, however, users can create more complex, dynamic collections.

Second, the concept of flexible collections allows users to put different types of information items (e.g., email messages, appointments, and web bookmarks) together into a single collection (Dourish et al., 1999; Karger and Quan, 2004). In this way, information organization is abstracted from the applications that produce and manage the items. Users can organize their items in a more logical and meaningful way, such as in terms of projects (Kaptelinin, 2003). As needed, users may group relevant files, email messages, web bookmarks, appointments, and to-do lists together with other types of information items. This principle reduces the need to maintain multiple hierarchies in various PIM tools, alleviating information fragmentation on a desktop.

2.2. Flat

While a hierarchical structure is pervasive in information management, organizing information items into a flat structure has recently gained popularity. In a flat structure, users assign tags or attributes to information items. These tags are used to group or retrieve the information items, providing associative access to the items (Dourish et al., 1999, 2000; Gifford et al., 1991; Gopal and Manber, 1999). This approach is now known as tagging.

Tagging provides a flexible way to organize information items. Users can classify an information item into multiple categories by assigning multiple tags to the item. Grouping and re-grouping information items can be done flexibly. For example, a bioinformatics article might be tagged “paper,” “bioinformatics,” “protein,” and “algorithms.” Having multiple tags, this article can be grouped with other documents by different categories: paper, bioinformatics, protein, and algorithms. Users can use any of these keywords to retrieve the paper. Compared to hierarchical approaches, tagging provides a lightweight mechanism for assigning multiple associations to an information item: users do not need to make shortcuts or copies of a document in order to put it in multiple categories.

The idea of tagging has existed for a long time (e.g., Dourish et al., 1999), but its recent applications in web-based systems along the lines of Web 2.0 have gained popularity (Golder and Huberman, 2006; Marlow et al., 2006). In these applications, users manage and assign tags to their shared collections on the Web. Examples include tools for organizing and sharing photos (http://www.flickr.com), web bookmarks (http://del.icio.us), and articles (http://www.citeulike.org).

Web-based tagging systems can be considered to be social, lightweight information retrieval systems. These systems share user-defined tags with all users and enable users to find other people who share the same items. Using this information, people can visit and browse the collections of other users who show similar interests (e.g., using the same tags or keeping the same information items). Browsing similarly tagged collections may lead to serendipitous discovery of useful information items.

Social tagging systems can also recommend tags to users based on common keywords used to tag shared items. Such recommendations allow users to see related keywords, and lessen the burden of creating categories for classifying information items. The recommendations result from a simple tag sharing, instead of from heavyweight reasoning mechanisms such as natural language analysis. In social tagging systems, users indirectly collaborate with one another to organize information items, utilizing the notion of the wisdom of crowds (Surowiecki, 2004).

Although tagging systems offer some advantages, they have shortcomings too. The freedom to associate multiple tags with an information item results in inconsistency in assigning tags. This inconsistency prevents users from retrieving all relevant items in a collection at once. Consider the following example. To ease retrieval, a user might tag research-related items with "research," "thesis," and "projects." However, this user does not use these tags consistently. Sometimes the user only uses "research" and "thesis." At other times, only "thesis" and "projects" are used. Consequently, when the user needs to retrieve all research-related items using the tag "research," the user will miss some of the relevant items (those tagged with "thesis" and "projects" only) due to such inconsistency.

Another problem with tagging approaches is that users are generally reluctant to annotate their information items extensively. For text files, attributes can be extracted automatically from the content of the files (Gifford et al., 1991). However, for other types of documents such as movies and pictures, automated tagging is challenging due to the difficulty in analyzing the content of such documents (see Smeulders et al. (2000) for a review of content-based image retrieval). In most cases, attribute assignment for non-text files still relies heavily on manual annotations. This can be problematic. When the number of items is still small, users may feel that assigning tags is not necessary, as they can find needed information items easily. As their collections grow, at some point, users will realize the usefulness of tags. At this point, however, their collections will have become too large to be annotated manually.

Creating tools for supporting manual annotations is one possible solution to this problem: for example, allowing users to annotate a group of items at once, to assign a set of attributes to every visited web page during a web-browsing session, and to give audio annotation, which is automatically transcribed to support text-based retrieval (Gemmell et al., 2002). In the case of shared collections, providing some incentives or rewards may be able to encourage users to tag shared information items as well.

Another approach to annotating information items is by using context analysis (Soules and Ganger, 2003). This approach is categorized into two classes: access-based and inter-file analysis. *Access-based analysis* uses existing information that is already available or given by users as annotations. For example, after submitting a query "personal information management," a user downloads a document from the Web and saves it in a directory "/research/pim/papers." It is reasonable to assume that the downloaded document is related to both the query and the directory name. Therefore, the document can be tagged "personal information management," "research," "pim," and "papers." *Inter-file analysis* monitors user access patterns to find out possible relationships between files. For example, if a user often opens files "project.doc" and "budget.doc" at the same time, then these files may be related or used in the same context. Therefore, it is logical to assign the files' attributes to each other. Such automated tagging can enrich user-defined tags, as it may detect useful patterns that otherwise go unnoticed by users.

2.3. Linear

In a linear structure, information items are arranged in a list based on certain order. The location of an item in the list is determined according to a particular attribute used to compare the item with other items. Examples of a linear structure include the order of words in dictionaries (alphabetical), entries in weblogs (chronological), and incoming messages in email inboxes (chronological). As long as users know the attribute of the information items they are looking for, having a sorted list allows them to traverse the list in a logical way, which is helpful in retrieval processes.

Among the common attributes used to sort a collection of information items, temporal attributes have received a lot of attention from PIM researchers. A chronological structure offers some advantages. First, time is a key retrieval cue in PIM (Cutrell et al., 2006a; Dumais et al., 2003; Graham et al., 2002; Malone, 1983; Rodden and Wood, 2003). People in general can roughly recall temporal attributes of information items, such as when an email message was received or when a picture was taken, and use this information to assist in finding needed information. A chronological structure is hence expected to improve information retrieval. Second, if information items are organized based on their temporal attributes, then information organization can be automated. This automation means that users do not have to spend a lot of efforts in categorizing documents. Third, a chronological structure may help in maintaining contextual information, as related information items are usually created or accessed about the same time, resulting in clusters of relevant items (Card et al., 1991).

A notable example of PIM systems using a linear structure is Lifestreams (Freeman and Gelernter, 1996). The system stores a user's document collection as a *lifestream*. A lifestream is the main stream of documents arranged in chronological order. It contains all information items that belong to the user. These items include archived documents, working documents, and possibly, future documents such as appointments and to-do lists, supporting the reminding function of PIM (Barreau and Nardi, 1995; Malone, 1983). Users do not have to name or categorize information items, as the system is responsible for placing the items in the lifestream and for assigning unique identifiers to them. However, as new information items arrive or are created constantly, a lifestream eventually becomes long and difficult to manage. Therefore, to facilitate information management, the system enables users to create substreams containing only subsets of their document collections. Substreams are dynamic views, created on demand to present query results. Within a substream, users may refine their queries further by creating other substreams; in this case, the system assumes logical conjunction between a query and its subquery. The system is also capable of summarizing information in a substream and presenting the result to the user. A sample application of this function is to summarize and chart time-series financial data.

Although time is an important retrieval cue, relying solely on it may hamper information retrieval. Particularly when the number of information items in a list is large, traversing the list to find a specific item is not easy. Moreover, there are other key retrieval cues as well in PIM, including visual, spatial, and social cues (Malone, 1983; Ravasio et al., 2004; Robertson et al., 1998; Whittaker et al., 2004). Therefore, instead of relying on a temporal attribute only, PIM systems should enable users to view a collection from multiple perspectives and to use as much contextual information as possible in the retrieval processes.

Another weakness of a linear structure is that it shows only a single dimension of an information collection at a time. When a collection is sorted alphabetically, the chronological order is lost, and vice versa. Moreover, unlike a hierarchical structure, a linear structure does not capture semantic relationships between information items explicitly. For example, the relationship between files “budget.doc” and “plan.doc,” which belong to the same project, is not obvious when the whole collection is arranged in a linear structure (e.g., in alphabetical/chronological order).

To overcome the limitations of a chronological structure while taking advantage of it, TimeSpace (Krishnan and Jones, 2005) allows users to create virtual workspaces for grouping related information items. These workspaces can be used to capture logical relation between information items. Items relating to the same activity can be grouped together in one workspace. TimeSpace can visualize user-defined workspaces along a time dimension to help users understand and keep track of their activities. Furthermore, the system complements the existing hierarchical file system instead of replacing it so that users can obtain the benefits of both organizational structures.

2.4. Spatial

Spatial approaches use locations as the main method of organizing information items. Spatial organization is pervasive in everyday life. People arrange things in their homes and offices so that they can find and use them easily. They separate important documents from unimportant ones on their desks. They keep frequently-used books in places that are easily accessible. In the digital world, a common example of spatial organization is computer desktop organization, where program shortcuts, file folders, and other items are arranged spatially to facilitate quick access to them (Ravasio et al., 2004).

A spatial structure maintains good visibility of information items. Spatially arranged documents can be scanned quickly, promoting recognition over recall and supporting quick access, finding, and reminding (Malone, 1983; Whittaker et al., 2004). A spatial approach to organizing web pages, for example, was shown to result in faster retrieval performance than the bookmarking facility in the Microsoft Internet Explorer (Robertson et al., 1998). Furthermore, the study participants were able to recall their spatial layouts of web pages a few months later with no significant differences in retrieval performance (Czerwinski et al., 1999).

Direct manipulation user interfaces (Shneiderman, 1997) serve well to support spatial organization of information items. Users just need to click and drag the items of interest to the desired locations. They see the results of their actions immediately and need neither create a hierarchy nor categorize the items explicitly, thereby reducing the burden of classifying information.

Spatial organization in computer systems, however, is limited by the size of computer monitors. With hundreds or thousands of information items, users simply cannot arrange all of the items spatially on their computer desktops without cluttering the desktops. Users want to keep their desktops tidy by filing away unneeded items, to avoid distraction and to allow them to concentrate on their work. Moreover, relying on pure spatial memory yields poor performance in information retrieval, particularly as the number of items increases (Jones and Dumais, 1986).

To deal with the limited size of screen real estate, Time-Machine Computing (TMC) records all state changes in its desktop (Rekimoto, 1999). TMC is a PIM system that extends the idea of Lifestreams (Freeman and Gelernter, 1996) by capturing both temporal and spatial attributes of documents. To do so, TMC provides a special desktop that allows users to arrange information items spatially without using folders as in a hierarchical structure. When users do not need some items any longer, or their desktops are full, they may remove unused items from the desktops by dragging them to a trash bin. Putting an item into the trash bin, however, does not delete the item permanently because TMC keeps track of any state changes on its desktop. When users need to access "deleted" items, they can set TMC's date back to a point before they put the items into the trash bin, and TMC will restore the state of the users' desktops at the specified time.

In addition to providing a special desktop to its users, TMC is capable of synchronizing the "current time" of multiple applications. The concept of current time can have different meanings for different applications. For email, the current time could refer to the arrival time of the currently selected message, whereas for a photo viewer, it could refer to the creation time of the currently displayed image. Assuming that each application has the concept of current time, TMC provides a mechanism for synchronizing the current time of its desktop and applications. For example, when a user selects an email message received a week ago, TMC can change the current time of the desktop and other applications accordingly: that is, TMC restores the state of the desktop back to last week. Thus, besides freeing users from categorizing information items into a hierarchy, TMC enables them to restore their workspaces at a specified time. This feature may help users recall contextual information about past activities.

2.5. Network

A network structure allows information items to be linked to one another arbitrarily. It does not impose any structural constraints on how users create links between information items. An example of a network structure on the global scale is the World Wide Web.

The ability to link information items arbitrarily characterizes both the strength and the weakness of a network structure. On one hand, this flexibility allows users to create information networks regardless of the types and physical locations of information items. For example, after receiving a document through email, a user can save the document into a file folder, and create a link between the document and the email message to preserve contextual information about the document, such as the sender, date received, and other information included in the email body (Bondarenko and Janssen, 2005). Users are free to wander around information networks by browsing and following links between information items. These features enable easy information sharing and transparent access to information.

On the other hand, since a network is less structured than other organizational structures, it is hard both to get an overview of an information network and to navigate through the network effectively: users are easily "lost" in hypertext systems (Cockburn and Jones, 1996; Komlodi et al., 2007; Olston and Chi, 2003). Furthermore, in network structures, broken links are a common error, which occurs when information items are removed from their previous locations on the servers.

Along with the development of the Semantic Web (Berners-Lee et al., 2001), there are ongoing efforts to improve PIM by using Semantic Web technologies (Decker et al., 2005) (see van Ossenbruggen et al. (2002) for an overview of Semantic Web technologies). The main principle is to allow users to describe their information items and create links between items using the Resource Description Framework (RDF) (Manola and Miller, 2004). Information items are seen as web resources and uniquely identified using the Uniform Resource Identifiers (URIs). For example, users can create descriptions about persons in their contact lists, identify each of them using a URI, and use the URI as a reference whenever necessary (e.g., using a URI to identify the author of a book or the sender of an email message). Based on this principle, users do not need to maintain copies of the same information item in multiple places; they just need to create a pointer to a URI representing an information item as needed.

The flexibility of RDF, which allows users to use arbitrary tags or terms to annotate information items, causes the vocabulary problem (Furnas et al., 1987): different users may use the same term to refer to different concepts, and vice versa. This problem must be resolved to enable software agents to process information meaningfully. Without a common understanding, software agents will not be able to communicate with one another properly. How can we address this vocabulary problem?

An approach to solving the vocabulary problem is to define or adopt an ontology (Gruber, 1995). An ontology contains formal definitions of terms and relationships among them in a certain domain, resulting in a set of vocabulary to discuss the domain. Based on an ontology, a set of inference rules can be defined to equip software agents with the ability to reason and derive new knowledge from the existing assertions in the ontology. Ontological approaches, however, only solve the vocabulary problem in limited domains because ontologies may not be compatible with one another. Furthermore, ontological approaches may not appeal to many users, as they increase cognitive load compared to other lightweight approaches to PIM (e.g., tagging).

The application of Semantic Web technologies in PIM enables the creation of semantically rich PIM tools termed Semantic Desktop systems (Decker and Frank, 2004; Huynh et al., 2002; Sauermann et al., 2005). Conceptually, Semantic Desktop systems are able to capture and describe both the structure and the semantics of an information network on a user's personal computer as metadata. Because this metadata can be "understood" by machines, this approach has potential to improve current practices in PIM. Information overload can be reduced by delegating well-defined tasks to software agents. Information sharing can become more meaningful and contextually rich, as relevant metadata can also be shared and described using common ontologies (Decker and Frank, 2004). Since the metadata is written in a standard format (RDF), Semantic Desktop systems can integrate shared metadata seamlessly into their local systems. In this way, the consumer of shared metadata does not have to do extra work, such as retyping the already available metadata. Furthermore, faceted metadata can be used to help users search and explore information collections (Yee et al., 2003).

Despite the potential, Semantic Desktop systems are still in early phase of development and must resolve various issues, such as user interface design, access control, semantic interoperability, and lack of powerful reasoning mechanisms, before they can deliver their full promises. Moreover, adopting an ontology to manage personal information spaces may be seen as a burden rather than a solution by users. Individual users have their own styles and preferences in managing personal information items. Some prefer to create elaborate organizational structures, while others simply put their information items in one folder (Abrams et al., 1998; Boardman and Sasse, 2004; Whittaker and Sidner, 1996). Personal contexts such as the purpose of a document affect how a user would organize the item (Barreau, 1995; Kwasnik, 1989, 1991). Thus, to be successful, Semantic Desktop frameworks must be able to accommodate individual purposes and be supported by tools that can automate and simplify the metadata creation processes as much as possible.

3. Implications for Design

As discussed in the previous section, each organizational structure has its own strengths and weaknesses. Therefore, replacing one structure with another rarely works well because it essentially replaces one's weaknesses with another's. Furthermore, no single structure can suit every user and every purpose or task. A hierarchical structure may be suitable for people who like to categorize information items regularly, but not for those who like to save everything in one big folder (e.g., "My Documents"). People dealing with time-series data may best use chronological approaches, whereas knowledge workers may prefer spatial organization to maintain the visibility of their documents (Kidd, 1994). Tagging may be suitable for managing general collections (e.g., shared web bookmarks), but when it comes to project management, people may prefer to create elaborate hierarchies to capture the structure of their information collections explicitly. PIM tool designers, therefore, should work toward augmenting existing structures instead of replacing one structure with another. In this section, we discuss some design recommendations to improve PIM tools.

3.1. Supporting Logical Organization and Task Management

Task management is an integral part of PIM. Unlike librarians, whose main task is to organize information items according to a standard classification scheme, users manage personal information items as a means to an end, usually to support their tasks. Thus, depending on its usage, the same information item may be categorized differently by different users. Functions pertinent to task management that need to be supported include reminding users of their to-do lists and appointments (Barreau and Nardi, 1995), helping them keep track of distributed tasks (Erickson et al., 2004), and facilitating contact management (Whittaker et al., 2004). To achieve these objectives, people use a variety of PIM tools, such as calendars (Blandford and Green, 2001; Kincaid et al., 1985), email clients (Whittaker and Sidner, 1996), and file systems (Jones et al., 2005). On one hand, using a specialized tool to manage a certain type of information items reduces the complexity of the tool. On the other hand, information items that are used in the same context are fragmented in various PIM tools, for example, information items relating to a project are stored in a file system, an email client, and a calendar (Karger and Jones, 2006; Ravasio et al., 2004; Whittaker et al., 2006).

To facilitate task management, PIM tools should support logical organization. The organization of information items should be abstracted both from the underlying storage structure (e.g., hierarchical, flat, linear) and from the tools that manage the items (e.g., calendars, file systems, email clients). Users should be able to group different types of information items in a way that is most meaningful to them. PIM tools should support organizational units that reflect higher-level user activities, for example in terms of tasks or projects (Bellotti et al., 2005; Kaptelinin, 2003; Krishnan and Jones, 2005). Multiple associations of an information item should be enabled in a lightweight manner without sacrificing the data integrity of the item or increasing the maintenance efforts (as in the case of maintaining multiple copies of a document in multiple places).

Interoperability rather than integration of PIM tools is more important to improve current practices in PIM. To enable logical organization of information items, developers may aim at delivering an integrated solution to PIM. In practice, however, people may prefer to use separate, lightweight tools for managing different types of information items. Reasons may vary from keeping the simplicity of the tools to avoiding unnecessary distraction while working on a task at hand. Even for managing the same type of information items (e.g., calendar entries, to-do lists), people use multiple tools for various reasons: for example, to separate personal items from work-related items or to ease scheduling of group meetings by using a shared calendar besides a personal calendar (Blandford and Green, 2001; Kincaid et al., 1985). To support this practice, PIM tools should enable users to access and synchronize entries in different tools.

The need for interoperability of PIM tools becomes more apparent when we consider user practice of using portable devices for task management. Related information items are fragmented not only across PIM tools, but also on various devices such as desktops, laptops, cell phones, and personal digital assistants (PDAs). People now need to create logical organization of information items across different PIM tools and devices. However, for some reasons, they may not want to synchronize data across these tools and devices all the time (Tungare and Pérez-Quiñones, 2008); what they need is ubiquitous and seamless access to information items. In this way, access to needed information is possible while allowing users to divide their information spaces for better management (e.g., personal vs. work-related spaces).

A way to support interoperability and logical organization is to use Semantic Web technologies (Berners-Lee et al., 2001; Sauermann et al., 2005): assigning a URI to every information item and using the URI as a reference when needed, akin to the way people use a Uniform Resource Locator to refer to a web resource. In this way, people can access and organize a URI across various PIM tools and devices, but only need to maintain the corresponding item in a single place, thereby preserving the data integrity of the item. Furthermore, since a URI is a globally unique identifier, it is possible to access an information item from anywhere, provided that the network infrastructure is available. The boundaries among PIM tools and among devices used to store and manage information items will eventually disappear. Due to the potential of the Semantic Web, we expect to see more applications of Semantic Web technologies in PIM.

3.2. Providing Multiple Visualizations

To take advantage of the existing organizational structures, we propose to treat them as *views* besides as methods to manage information items. Since changing user practices abruptly may lead to resistance to using a new tool, PIM tools should add more features on top of the predominant hierarchical structure. Thus, in addition to using hierarchical folders, users should be able to tag information items, arrange them spatially (as in the way they place items on computer desktops), and create logical organizational units or arbitrary links between items. Then PIM tools should provide several options on how users want to view their collections. For example, users should be able to generate a chronological view of their collections, with options whether to apply it to a particular type of items (e.g., email messages only) or across several types of information items (e.g., email messages, web bookmarks, to-do lists, files). Other views such as spatial and network views should also be supported. These views allow users to see an information collection from various perspectives and to use different kinds of contextual information as retrieval cues. For example, users may forget the filename of a document they are looking for, but remember that they opened the document yesterday or that they linked the document with another document at hand. If users have access to various views, they will be able to retrieve the needed document by generating a chronological or network view, and then using the last accessed date as a retrieval cue or following a link from the document at hand.

To implement views, developers can use information visualization techniques and follow heuristic guidelines such as “overview first, zoom and filter, then details-on-demand" (Shneiderman, 1996). One of the strengths of visualization is that it can present multiple attributes of an information collection simultaneously so that users can explore, compare, and analyze the collection from various dimensions. Emerging patterns in visualization can be recognized easily by human vision (Ware, 2000). Researchers have also developed techniques for visualizing different types of data sets, including hierarchical data sets (Bederson et al., 2002), graphs (Herman et al., 2000), hyperlink or network structures (Zhang and Nguyen, 2005), and temporal data (Plaisant et al., 1996). These techniques may serve as a basis for implementing various views to support PIM.

3.3. Utilizing Past and Planned Interaction Records

One of the advantages of digital information spaces over physical spaces is that activities that users do in digital spaces leave electronic traces. These traces have many potential applications in PIM. For example, access patterns to information items can be monitored and visualized to give overviews of users' activities over time (Kaptelinin, 2003; Krishnan and Jones, 2005). Email archives can be analyzed to reveal social network structures and interaction patterns to help users recall contextual information about their past work (Fisher and Dourish, 2004). Analysis of email archives can also suggest important contacts, which can then be organized spatially for social reminding and social data mining (Whittaker et al., 2004). Search systems may unobtrusively monitor and keep a user's web browsing history; augment it with metadata, such as the last access time to particular web pages or keywords used to find particular articles; and then allow the user to re-find information items based on these associative local cues (Cutrell et al., 2006b; Komlodi et al., 2007). In principle, PIM tools should utilize electronic traces to provide personalized environments for users.

Besides utilizing interaction histories, PIM tools can also use future or planned interaction records, which often are available in calendars, to provide better support for users. Common features of electronic calendars include reminding users of their appointments, giving a warning when time conflicts occur, and suggesting possible time slots for group meetings. The applications of electronic calendars, however, should not be limited to time management. Taking a holistic approach, a PIM system may combine analysis of calendar entries and information items in other tools to facilitate information and task management. Consider the following scenario. A user enters a calendar event describing a plan to attend a conference. The calendar entry contains information such as the date, name, and location of the conference. It is likely that relevant documents also exist in the user's file and email folders. When sending a reminder about the event, a PIM system may also include links to potentially relevant documents (e.g., a link to a file folder that has a name similar to the calendar event). Seeing such information may remind the user to bring necessary documents to the conference or to finish an outstanding task such as preparing slides for the conference presentation. At a later date, when the user uploads photos taken during the conference period, the PIM system may suggest tags based on information in the calendar entry.

In using interaction records, PIM tools must consider that users' work contexts change from time to time (Kaptelinin, 2003). While working on a document, users sometimes access unrelated documents, such as email messages or news websites, to take a break. They sometimes get interruptions that require them to look up some information. Such activities leave electronic traces that seem to be relevant with the users' current work contexts, which, in fact, are misleading. If the systems do not consider such context switching, they will end up linking irrelevant information items.

Even for a planned activity listed in a calendar, the actual activity may not exactly match the description in the calendar: for example, a meeting may last longer than what is scheduled, a task may require additional time to finish, and an empty slot in a user's calendar does not automatically imply that the user has free time during that period (Blandford and Green, 2001). Therefore, PIM tools should not rely too much on past and future interaction records, and should facilitate users to configure their workspaces manually.

3.4. Integrating Social Contexts into PIM

Although often seen as a personal activity, PIM fundamentally is both personal and social activities (Erickson, 2006). It is personal, as users organize personal information spaces using their own ways, mainly to serve their tasks. In practice, however, personal uses of information items usually fall into a larger, social context. For example, people who work on the same project need to manage and use shared materials. Upon encountering relevant information, people share it with colleagues, friends, or family members (Marshall and Bly, 2005). Moreover, with the widely available network connectivity, computers are no longer isolated devices, but have become an access point to information and social networks, connecting people with one another. As a result, groupware applications have been developed to support group work. Online communities emerge as a place to discuss various topics of interests and to support collaborative work. And PIM has also started to move toward a lightweight, collaborative activity, as exemplified by the popularity of peer-to-peer file-sharing networks (Androutsellis-Theotokis and Spinellis, 2004), weblogs (Nardi et al., 2004), and social tagging systems (Golder and Huberman, 2006; Marlow et al., 2006).

Social elements of PIM are currently supported by computer-mediated communication applications, such as email and instant messenger. People use these applications to facilitate collaborative processes and to maintain shared awareness with their colleagues. Although supporting collaborative work has not become a main concern in PIM, we expect to see more integration between computer-supported cooperative work (CSCW) applications and PIM tools in the future. With advances in hardware (e.g., Microsoft Surface – http://www.microsoft.com/surface/), people will eventually be able to integrate the management of tangible and digital documents, to transfer data between devices easily, and to collaborate with other people seamlessly. The boundary between personal and social information spaces will become *fluid*: whenever necessary, a personal space should be able to turn into a collaborative space instantly while still protecting other non-shared documents from unauthorized access. The integration of CSCW applications into PIM systems will enable lightweight, informal, and opportunistic collaboration (Gutwin et al., 2005), which is typical in the real world.

4. Research Aspects of Personal Information Management

To give further ideas for future work, we propose a faceted classification scheme to describe research aspects of PIM. The classification scheme consists of four facets: the type of information items, the number of users, the number of application domains, and the number of devices (Figure 1). The research aspects of each facet are not mutually exclusive.

The type of information items:
Physical – Digital

The number of users:
Single – Multi-user

The number of application domains:
Single – Multi-domain

The number of devices:
Single – Multi-device

Figure 1: A faceted classification scheme of research aspects of PIM.

The first facet classifies PIM research according to the type of information items being studied. This facet has two categories: physical and digital. Research that examines management of paper documents (Whittaker and Hirschberg, 2001) or physical spaces such as offices (Malone, 1983) falls into the physical category, whereas studies concerning PIM in digital spaces such as computer desktop organization (Ravasio et al., 2004) belong to the digital category.

The second facet is concerned with whether a study focuses on single- or multi-user applications. To date, the majority of PIM research studies individual user practices of managing information, and hence falls into the single user category. This is not surprising, as most PIM tools are designed and intended to be single-user applications. However, there is a trend towards developing PIM tools as multi-user applications, such as shared calendars and social bookmarking systems (e.g., http://del.icio.us and http://www.citeulike.org). Thus, research that focuses on such applications or collaborative uses of PIM tools belongs to the multi-user category (e.g., Golder and Huberman, 2006; Marlow et al., 2006).

The third facet categorizes research based on the number of application domains in the study. For example, studies of management of email messages (Whittaker and Sidner, 1996) or web bookmarks (Abrams et al., 1998) are classified into the single domain category. The multi-domain category is reserved for research involving multiple PIM tools, such as a cross-tool study of management of files, email messages, and web bookmarks (Boardman and Sasse, 2004), or a tool that supports multiple aspects of PIM, such as an email client with integrated task management facility (Bellotti et al., 2005).

Finally, the last facet describes whether PIM research focuses on a single device (e.g., computer desktops) or multiple devices (e.g., computer desktops, cell phones, and PDAs). With advances in portable devices, PIM is becoming increasingly distributed. For example, people maintain multiple contact lists on computer desktops, cell phones, and PDAs. While traveling they may need to access information items on their desktops using PDAs. These practices make the task of designing PIM tools more challenging, as developers have to consider not only interoperability among applications on a single device, but also interoperability among applications on multiple devices.

This classification scheme provides a framework for researchers to experiment with different ideas for designing and developing novel PIM tools. Looking at the facets in the classification scheme, researchers may pose questions such as how to facilitate management of documents in a collaborative and multi-application context, or in a collaborative context and on multiple devices. While "classic" research usually addresses single dimensions of the facets, current research has started to address the right side of the facets. We can anticipate future work to address combination of dimensions and problems in this classification scheme. For example, developing intuitive, visual metaphors for managing documents collaboratively across different devices would be an interesting direction of research.

5. Concluding Remarks

PIM research is moving towards a holistic approach. This trend results from the user practices of managing personal information spaces. Users manage information items to support their tasks. Task management usually involves various types of information items (e.g., email messages, files, calendar entries), which are managed by different PIM tools and sometimes stored on multiple devices. Considering the nature of PIM, researchers should strive to provide a holistic solution to PIM, especially to support larger contexts of PIM such as task and time management. This approach requires fluid boundaries among PIM tools both on a single device and on multiple devices. A way to achieve this interoperability is to have a unified representation of information items (Karger and Jones, 2006).

Since task management is also a social process, it is desirable to have a fluid boundary between personal and social spaces too. The most important purpose of PIM tools is to support individual user needs. However, users sometimes need to share information items or communicate with other people to facilitate collaborative work. If PIM tools enable users to create collaborative workspaces, users will be able to manage their tasks better. Eventually, there will be integration of CSCW applications into PIM tools.

In summary, we reviewed and categorized existing approaches to PIM based on their organizational structures. We identified the strengths and weaknesses of each structure and discussed their implications for the design of PIM tools. Our design recommendations for PIM tools are as follows: (1) supporting logical organizational units; (2) providing multiple visualizations of personal information spaces; (3) utilizing past and planned interaction records to enhance PIM; and (4) integrating social contexts into PIM. In addition, we proposed a faceted classification scheme to help researchers generate ideas for developing new PIM tools.

Acknowledgements

We are grateful for support from Natural Sciences and Engineering Research Council of Canada through the LORNET Research Network.

References

Abrams, D., Baecker, R., Chignell, M. (1998) "Information Archiving with Bookmarks: Personal Web Space Construction and Organization". In *Proceedings of the SIGCHI conference on human factors in computing systems*, pp. 41–48

Androutsellis-Theotokis, S., Spinellis, D. (2004) “A Survey of Peer-to-Peer Content Distribution Technologies”. *ACM Computing Surveys*, Vol. 36, No. 4, 335–371

Barreau, D. (1995) “Context as a Factor in Personal Information Management Systems”. *Journal of the American Society for Information Science*, Vol. 46, No. 5, 327–339

Barreau, D., Nardi, B. (1995) “Finding and Reminding: File Organization from the Desktop”. *SIGCHI Bulletin*, Vol. 27, No. 3, 39–43

Bederson, B.B., Shneiderman, B., Wattenberg, M. (2002) “Ordered and Quantum Treemaps: Making Effective Use of 2D Space to Display Hierarchies”. *ACM Transactions on Graphics*, Vol. 21, No. 4, 833–854

Bellotti, V., Ducheneaut, N., Howard, M., Smith, I., Grinter, R.E. (2005) “Quality versus Quantity: E-mail-Centric Task Management and Its Relation with Overload”. *Human-Computer Interaction*, Vol. 20, No. 1-2, 89–138

Bergman, O., Beyth-Marom, R., Nachmias, R. (2006) “The Project Fragmentation Problem in Personal Information Management”. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pp. 271–274

Berners-Lee, T., Hendler, J., Lassila, O. (2001) “The Semantic Web”. *Scientific American*, Vol. 284, No. 5, 34–43

Blandford, A.E., Green, T.R.G. (2001) “Group and Individual Time Management Tools: What You Get is Not What You Need”. *Personal and Ubiquitous Computing*, Vol. 5, No. 4, 213–230

Boardman, R., Sasse, M.A. (2004) ““Stuff goes into the computer and doesn’t come out”: A Cross-Tool Study of Personal Information Management”. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pp. 583–590

Boardman, R., Spence, R., Sasse, M.A. (2003) “Too Many Hierarchies? The Daily Struggle for Control of the Workspace”. In *Proceedings of HCI international*, pp. 616–620

Bondarenko, O., Janssen, R. (2005) “Documents at Hand: Learning from Paper to Improve Digital Technologies”. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pp. 121–130

Card, S.K., Robertson, G.G., Mackinlay, J.D. (1991) “The Information Visualizer, an Information Workspace”. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pp. 181–188

Cockburn, A., Jones, S. (1996) “Which Way Now? Analysing and Easing Inadequacies in WWW Navigation”. *International Journal of Human-Computer Studies*, Vol. 45, No. 1, 105–129

Czerwinski, M., van Dantzich, M., Robertson, G., Hoffman, H. (1999) “The Contribution of Thumbnail Image, Mouse-Over Text and Spatial Location Memory to Web Page Retrieval in 3D”. In *Proceedings of Interact ’99*, pp. 163–170

Cutrell, E., Dumais, S.T., Teevan, J. (2006a) “Searching to Eliminate Personal Information Management”. *Communications of the ACM*, Vol. 49, No. 1, 58–64

Cutrell, E., Robbins, D.C., Dumais, S.T., Sarin, R. (2006b) “Fast, Flexible Filtering with Phlat– Personal Search and Organization Made Easy”. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pp. 261–270

Decker, S., Park, J., Quan, D., Sauermann, L. (Eds.) (2005) *Proceedings of the workshop on the semantic desktop*.

Decker, S., Frank, M. (2004) “The Networked Semantic Desktop”. In *Proceedings of the WWW 2004 workshop on application design, development and implementation issues in the semantic web*.

Dourish, P., Edwards, W.K., LaMarca, A., Lamping, J., Petersen, K., Salisbury, M., Terry, D.B., Thornton, J. (2000) “Extending Document Management Systems with User-Specific Active Properties”. *ACM Transactions on Information Systems*, Vol. 18, No. 2, 140–170

Dourish, P., Edwards, W.K., LaMarca, A., Salisbury, M. (1999) “Presto: An Experimental Architecture for Fluid Interactive Document Spaces”. *ACM Transactions on Computer-Human Interaction*, Vol. 6, No. 2, 133–161

Dumais, S., Cutrell, E., Cadiz, J.J., Jancke, G., Sarin, R., Robbins, D.C. (2003) “Stuff I’ve Seen: A System for Personal Information Retrieval and Re-Use”. In *Proceedings of the 26th annual international ACM SIGIR conference on research and development in information retrieval*, pp. 72–79

Erickson, T. (2006) “From PIM to GIM: Personal Information Management in Group Contexts”. *Communications of the ACM*, Vol. 49, No. 1, 74–75

Erickson, T., Huang, W., Danis, C., Kellogg, W.A. (2004) “A Social Proxy for Distributed Tasks: Design and Evaluation of a Working Prototype”. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pp. 559–566

Fisher, D., Dourish, P. (2004) “Social and Temporal Structures in Everyday Collaboration”. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pp. 551–558

Freeman, E., Gelernter, D. (1996) “Lifestreams: A Storage Model for Personal Data”. *SIGMOD Record*, Vol. 25, No. 1, 80–86

Furnas, G.W., Landauer, T.K., Gomez, L.M., Dumais, S.T. (1987) “The Vocabulary Problem in Human-System Communication”. *Communications of the ACM*, Vol. 30, No. 11, 964–971

Gemmell, J., Bell, G., Lueder, R., Drucker, S., Wong, C. (2002) “MyLifeBits: Fulfilling the Memex Vision”. In *Proceedings of the 10th ACM international conference on multimedia*, pp. 235–238

Gifford, D.K., Jouvelot, P., Sheldon, M.A., O’Toole, J.W. (1991) “Semantic File Systems”. In *Proceedings of the 13th ACM symposium on operating systems principles*, pp. 16–25

Golder, S.A., Huberman, B.A. (2006) “Usage Patterns of Collaborative Tagging Systems”. *Journal of Information Science*, Vol. 32, No. 2, 198–208

Gopal, B., Manber, U. (1999) “Integrating Content-Based Access Mechanisms with Hierarchical File Systems”. In *Proceedings of the 3rd symposium on operating systems design and implementation*, pp. 265–278

Graham, A., Garcia-Molina, H., Paepcke, A., Winograd, T. (2002) “Time as Essence for Photo Browsing through Personal Digital Libraries”. In *Proceedings of the 2nd ACM/IEEE-CS joint conference on digital libraries*, pp. 326–335

Gruber, T.R. (1995) “Toward Principles for the Design of Ontologies Used for Knowledge Sharing”. *International Journal of Human-Computer Studies*, Vol. 43, No. 5-6, 907–928

Gutwin, C., Greenberg, S., Blum, R., Dyck, J. (2005) “Supporting Informal Collaboration in Shared-Workspace Groupware”. Technical report HCI-TR-05-01, Department of Computer Science, University of Saskatchewan. http://hci.usask.ca/publications/2005/hci-tr-05-01.pdf

Herman, I., Melançon, G., Marshall, M.S. (2000) “Graph Visualization and Navigation in Information Visualization: A Survey”. *IEEE Transactions on Visualization and Computer Graphics*, Vol. 6, No. 1, 24–43

Huynh, D., Karger, D.R., Quan, D. (2002) “Haystack: A Platform for Creating, Organizing and Visualizing Information Using RDF”. In *Proceedings of the WWW 2002 semantic web workshop*.

Jones, W., Phuwanartnurak, A.J., Gill, R., Bruce, H. (2005) “Don’t Take My Folders Away! Organizing Personal Information to Get Things Done”. In *CHI ’05 extended abstracts on human factors in computing systems*, pp. 1505–1508

Jones, W.P., Dumais, S.T. (1986) “The Spatial Metaphor for User Interfaces: Experimental Tests of Reference by Location versus Name”. *ACM Transactions on Information Systems*, Vol. 4, No. 1, 42–63

Kaptelinin, V. (2003) “UMEA: Translating Interaction Histories into Project Contexts”. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pp. 353–360

Karger, D.R., Jones, W. (2006) “Data Unification in Personal Information Management”. *Communications of the ACM*, Vol. 49, No. 1, 77–82

Karger, D.R., Quan, D. (2004) “Collections: Flexible, Essential Tools for Information Management”. In *CHI ’04 extended abstracts on human factors in computing systems*, pp. 1159–1162

Kidd, A. (1994) “The Marks are on the Knowledge Worker”. In *Proceedings of the SIGCHI conference on human factors in computing systems*, pp. 186–191

Kincaid, C.M., Dupont, P.B., Kaye, A.R. (1985) “Electronic Calendars in the Office: An Assessment of User Needs and Current Technology”. *ACM Transactions on Information Systems*, Vol. 3, No. 1, 89–102

Komlodi, A., Marchionini, G., Soergel, D. (2007) “Search History Support for Finding and Using Information: User Interface Design Recommendations from a User Study”. *Information Processing & Management*, Vol. 43, No. 1, 10–29

Krishnan, A., Jones, S. (2005) “TimeSpace: Activity-Based Temporal Visualisation of Personal Information Spaces”. *Personal and Ubiquitous Computing*, Vol. 9, No. 1, 46–65

Kwasnik, B. (1991) “The Importance of Factors that are not Document Attributes in the Organisation of Personal Documents”. *Journal of Documentation*, Vol. 47, No. 4, 389–398

Kwasnik, B. (1989) “How a Personal Document’s Intended Use or Purpose Affects Its Classification in an Office”. In *Proceedings of the 12th annual international ACM SIGIR conference on research and development in information retrieval*, pp. 207–210

Lansdale, M. (1988) “The Psychology of Personal Information Management”. *Applied Ergonomics*, Vol. 19, No. 1, 55–66

Malone, T.W. (1983) “How do People Organize Their Desks? Implications for the Design of Office Information Systems”. *ACM Transactions on Information Systems*, Vol. 1, No. 1, 99–112

Mander, R., Salomon, G., Wong, Y.Y. (1992) "A 'Pile' Metaphor for Supporting Casual Organization of Information". In *Proceedings of the SIGCHI conference on human factors in computing systems*, pp. 627–634

Manola, F., Miller, E. (Eds) (2004) "RDF Primer". World Wide Web Consortium. http://www.w3.org/TR/rdf-primer/

Marlow, C., Naaman, M., boyd, d., Davis, M. (2006) "HT06, Tagging Paper, Taxonomy, Flickr, Academic Article, to Read". In *Proceedings of the 17th conference on hypertext and hypermedia*, pp. 31–40

Marshall, C.C., Bly, S. (2005) "Saving and Using Encountered Information: Implications for Electronic Periodicals". In *Proceedings of the SIGCHI conference on human factors in computing systems*, pp. 111–120

Nardi, B., Schiano, D.J., Gumbrecht, M., Swartz, L. (2004) "Why We Blog". *Communications of the ACM*, Vol. 47, No. 12, 41–46

Olston, C. and Chi, E.H. (2003) "ScentTrails: Integrating Browsing and Searching on the Web". *ACM Transactions on Computer-Human Interaction*, Vol. 10, No. 3, 177–197

Plaisant, C., Milash, B., Rose, A., Widoff, S., Shneiderman, B. (1996) "LifeLines: Visualizing Personal Histories". In *Proceedings of the SIGCHI conference on human factors in computing systems*, pp. 221–227

Ravasio, P., Schär, S.G., Krueger, H. (2004) "In Pursuit of Desktop Evolution: User Problems and Practices with Modern Desktop Systems". *ACM Transactions on Computer-Human Interaction*, Vol. 11, No. 2, 156–180

Rekimoto, J. (1999) "Time-Machine Computing: A Time-Centric Approach for the Information Environment". In *Proceedings of the 12th annual ACM symposium on user interface software and technology*, pp. 45–54

Robertson, G., Czerwinski, M., Larson, K., Robbins, D.C., Thiel, D., van Dantzich, M. (1998) "Data Mountain: Using Spatial Memory for Document Management". In *Proceedings of the 11th annual ACM symposium on user interface software and technology*, pp. 153–162

Rodden, K., Wood, K.R. (2003) "How do People Manage Their Digital Photographs?" In *Proceedings of the SIGCHI conference on human factors in computing systems*, pp. 409–416

Sauermann, L., Bernardi, A., Dengel, A. (2005) "Overview and Outlook on the Semantic Desktop". In *Proceedings of the workshop on the semantic desktop*, pp. 1–18

Shneiderman, B. (1997) "Direct Manipulation for Comprehensible, Predictable and Controllable User Interfaces". In *Proceedings of the 2nd international conference on intelligent user interfaces*, pp. 33–39

Shneiderman, B. (1996) "The Eyes Have It: A Task by Data Type Taxonomy for Information Visualizations". In *Proceedings of the IEEE symposium on visual languages*, pp. 336–343

Smeulders, A.W.M., Worring, M., Santini, S., Gupta, A., Jain, R. (2000) "Content-Based Image Retrieval at the End of the Early Years". *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Vol. 22, No. 12, 1349–1380

Soules, C.A.N., Ganger, G.R. (2003) "Why Can't I Find My Files? New Methods for Automating Attribute Assignment". In *Proceedings of the 9th workshop on hot topics in operating systems*, pp. 115–120

Surowiecki, J. (2004) *The Wisdom of Crowds: Why the Many are Smarter than the Few and How Collective Wisdom Shapes Business, Economies, Societies, and Nations* (Doubleday)

Teevan, J., Jones, W., Bederson, B.B. (2006) "Personal Information Management". *Communications of the ACM*, Vol. 49, No. 1, 40–43

Tungare, M., Pérez-Quiñones, M. (2008) "Thinking Outside the (Beige) Box: Personal Information Management Beyond the Desktop". In *Proceedings of the CHI workshop on personal information management*, http://pim2008.ethz.ch/papers/pim2008-tungare-etal.pdf

van Ossenbruggen, J., Hardman, L., Rutledge, L. (2002) "Hypermedia and the Semantic Web: A Research Agenda". *Journal of Digital Information*, Vol. 3, No. 1, http://journals.tdl.org/jodi/article/view/jodi-61/77

Ware, C. (2000) *Information Visualization: Perception for Design* (Morgan Kaufmann Publishers)

Whittaker, S., Bellotti, V., Gwizdka, J. (2006) "Email in Personal Information Management". *Communications of the ACM*, Vol. 49, No. 1, 68–73

Whittaker, S., Jones, Q., Nardi, B., Creech, M., Terveen, L., Isaacs, E., Hainsworth, J. (2004) "ContactMap: Organizing Communication in a Social Desktop". *ACM Transactions on Computer-Human Interaction*, Vol. 11, No. 4, 445–471

Whittaker, S., Hirschberg, J. (2001) "The Character, Value, and Management of Personal Paper Archives", *ACM Transactions on Computer-Human Interaction*, Vol. 8, No. 2, 150–170

Whittaker, S., Sidner, C. (1996) "Email Overload: Exploring Personal Information Management of Email". In *Proceedings of the SIGCHI conference on human factors in computing systems*, pp. 276–283

Yee, K.P., Swearingen, K., Li, K., Hearst, M. (2003) "Faceted Metadata for Image Search and Browsing". In *Proceedings of the SIGCHI conference on human factors in computing systems*, pp. 401–408

Zhang, J., Nguyen, T. (2005) "WebStar: A Visualization Model for Hyperlink Structures". *Information Processing & Management*, Vol. 41, No. 4, 1003–1018