Human and Social Aspects of Decentralized Knowledge Communities

Indratmo and Julita Vassileva

Department of Computer Science, University of Saskatchewan
Saskatoon, SK S7N 5C9, Canada
j.indratmo@usask.ca, jiv@cs.usask.ca

Abstract. To design an infrastructure for knowledge communities, we need both technical expertise and an understanding of human and social aspects of communities. Technologies for implementing such infrastructures are often available. However, there is no clear, proven procedure for building successful communities. In this paper, we review research literature concerning user practices and social aspects of information and knowledge management. Based on this review, we propose preliminary design criteria for Semantic Desktop systems.

1 Introduction

Designing an infrastructure for knowledge communities requires not only technical expertise, but also a proper understanding of the human and social aspects of communities. Understanding these aspects is necessary because the success of communities often depends on subtle design issues [1, 2].

The structure and dynamics of a community emerge from local interactions of its members. These interactions cannot be controlled to produce specific collective behaviour, as each community member is autonomous and interdependent. However, we can support communities by designing an infrastructure that fosters characteristics of successful communities, such as a high level of participation, contribution, and cooperation among community members.

Recently, a vision [3] has emerged of building peer-to-peer (P2P), semantically rich knowledge communities. The basic idea is to apply Semantic Web principles [4] to personal information management (PIM), resulting in Semantic Desktop systems (e.g., [5, 6]). Then these systems are interconnected using a P2P protocol, providing an infrastructure for online communities. Some of the technical aspects of such infrastructures have been discussed elsewhere (e.g., [3, 5–7]). However, the discussions so far have overlooked the existing user practices in PIM and social aspects of communities. Unless the new infrastructures accommodate these practices, they will not be adopted by users, and their applications will be severely limited.

In this paper, we review previous research on personal information management and social aspects of knowledge management, so that lessons from many user studies about PIM and online communities can be used for implementing the vision of P2P knowledge communities [3]. The paper consists of five main parts: (1) user practices in PIM, (2) problems in information management, (3) various approaches to solving the problems, (4) design criteria for Semantic Desktop systems, and (5) social aspects of knowledge communities.

In the first three parts, we give background information about PIM—activities in organizing and managing information for personal use. The discussion focuses on human factors in PIM, which then become the basis for the development of criteria for designing Semantic Desktop systems. In the last part, we discuss social aspects of communities, such as the role of social networks in information and knowledge management, and factors that motivate users to contribute to their communities.

2 User Practices in Personal Information Management

This section discusses existing research on user practices in personal information management. First, it outlines common purposes of PIM, then discusses different types of filing behaviours, and finally highlights the importance of various cues in information management.

2.1 Basic Purposes: Finding and Reminding

People manage their documents not only to make it easier to find them later, but also to remind themselves of their tasks [8, 9]. To make finding documents easier, people arrange the documents according to some logical order. To achieve the reminding function, they make relevant documents more visible than others. For example, putting a document on a computer desktop can remind the user to work on the document.

To support finding and reminding, people organize their documents using "files" and "piles" [8, 10]. Files are organized document collections and usually contain archived information. Piles, on the other hand, are disorganized document collections and typically contain work-in-progress documents. A pile can function as a reminder and a temporary organizational unit. Although not organized explicitly, elements of a pile may naturally follow reverse chronological order [8].

When searching for specific information on their personal computer, users prefer to browse their directories manually than to use a search tool [9, 11, 12]. This preference may result from users' familiarity with their personal workspace: they feel that they know where they have saved a document. Users tend to use a search tool only when they cannot find a document manually, or when they search information that is available outside their personal workspace, such as on a shared file server or on the Web.

2.2 Filing Behaviour

A common activity in information management is to file documents; that is, to place documents in an appropriate folder or category.

People's behaviour in organizing their documents ranges from "pilers" to "filers" [12–14]. Pilers are users who do not categorize their documents. An example is an email user who leaves all incoming messages in the inbox. Filers, on the other hand, are those who use folders extensively; they create many folders and classify their documents into various folders on a daily basis.

Personality traits and job types affect people's filing behaviour. Some people tend to be more organized than others. People whose tasks are structured and routine tend to file their documents more frequently than those whose tasks are unstructured [8]. For knowledge workers [15], for example, filing information is often not so important, as their work relies heavily on creativity and knowledge that is in their mind. Thus, knowledge workers focus on absorbing information contained in documents, rather than relying on them.

Researchers have observed filing behaviour in various applications, such as email [12, 14], Web bookmarks [12, 13], and file systems [12]. The classification of filing behaviour in these studies, however, is not absolute. People's behaviour is complex, so it is possible to refine the suggested classification schemes. Boardman and Sasse [12] observe that users in practice use multiple strategies to manage emails and Web bookmarks. Specifically, some email users file important messages immediately (frequent filers), but leave other messages in the inbox and organize them occasionally (spring cleaners). Similarly, Internet users organize some of their new Web bookmarks at the creation time (creation-time filers), but leave others as they are (no-filers).

The research discussed above studies how frequently and when people file their documents, but it does not discuss how people group their documents. What factors affect classificatory decisions?

Documents' intended use or purpose appears to be a strong factor that affects classificatory decisions [16]. People tend to group documents based on the task related to the documents, such as grouping all materials for teaching a course Computer Science 101. Basic properties of documents—their size, type, or author—are less influential than their intended use. In organizing photos [17], however, people typically use temporal information, such as events or date, as the main factor in making classificatory decisions.

2.3 The Importance of Context and Cues

People use various cues to maintain contextual information of their documents. A business card attached on a paper may serve as a reminder to send the paper to a particular person. A date written on top of a document may remind the user of a project's deadline. A red folder may indicate the urgency of documents within it. These cues, however, can be interpreted and used in different ways. Therefore, knowing the context in which such cues are being used is crucial.

Context is important for a proper understanding of information and in the recall process [18]. A classic example, as discussed in [18], is a chess study [19]. This study suggests that chess experts can reconstruct chess positions from real play well because the positions are meaningful to them, which helps in recall. When the positions are randomized, their ability to recall declines significantly.

In the context of information management, associating a document with various cues—visual, spatial, chronological, or contextual—can improve people's ability to find the document because humans are good at remembering the appearance, essence, and context of a document. Without such cues, people have to remember details, such as filenames, which they are not good at.

Spatial locations of documents also carry semantic information to the user. In office organization, people keep documents that are more recent or urgent closer to the centre of their working area [8, 20]. In desktop organization, computer users arrange their screen systematically [11]: they group program shortcuts on a specific region, put related documents close to each other, and try to arrange their screen's layout symmetrically. Such spatial organizations can help users locate frequently used programs and documents on their computer.

3 Problems in Information Management

Current systems for PIM are far from being perfect. Studies report four areas where improvement is necessary: supporting information organization better, maintaining context and retrieval cues, dealing with the problem of information overload, and supporting interoperability between individuals or organizations.

3.1 Filing

Filing information is difficult. It has been observed in various studies, including office organization [8, 10], email management [14], and desktop organization [11]. In these studies, users state that filing is a heavyweight cognitive activity: they cannot make classificatory decisions easily. When filing a document, users try to put the document into a category in which they expect to remember it for later retrieval. Because there are many ways to organize documents, as discussed in section 2.2, a classification scheme that is currently appropriate may become unsuitable as the user requirements or goals change over time.

The currently predominant hierarchical structure in information management makes filing information difficult and inflexible. In the physical world, people are limited to having to put a document into one specific folder, unless they make a copy of the document, which then can be put in a different folder. Unfortunately, most file systems and computer applications impose the same restriction: they do not allow users to classify a document into multiple categories (e.g., folders or directories). Although users can alleviate this problem by making copies of documents or links to files, they normally do not do this [10].

Hierarchies also restrict users from capturing multiple semantics of a document. Where should a user save a paper about the Semantic Web and ontology, in a “Semantic Web” or an “Ontology” directory? Filing the paper in either of these categories causes loss of information about the document's content. And it complicates the retrieval process, as the user has to remember exactly the location of the paper. Although the user could create a “Semantic Web and Ontology” directory, it is uncommon to have a category with multiple semantics.

The ambiguity of natural languages also contributes to the difficulty of filing information. A word can refer to different things. A category “Networks” may refer to network protocols, Internet applications, or even network marketing. Thus, it is difficult to choose good names for semantic categories. In addition, it is hard later to remember the intended semantics of the categories, as the mind-set of the user may change in the meantime. Often a category created with a specific semantics “evolves” due to filing of documents that do not exactly fit in it, because users are reluctant to create a new category, since this may require reorganization of the previously classified documents.

3.2 Maintaining Context and Retrieval Cues

Computer systems offer limited supports for maintaining context and retrieval cues in information management. Visual, contextual, and spatial cues are important for finding and reminding [18]. In the physical world, tangible documents have rich features, such as colour, size and thickness, which can serve as retrieval cues. For example, when searching for a book, people may forget the book’s title, but they usually can still remember the book’s appearance. Therefore, they can limit their search based on these cues. Unfortunately, many of these cues are lost in the digital world. Instead of maintaining visual retrieval cues, computer systems display details of files, such as their name, type, and size, which are less useful for retrieval cues. Furthermore, there are only limited possibilities for users to arrange the spatial layout of their document collections.

Filing documents removes some contextual information of the documents. Suppose that in an office, a user keeps urgent documents in a red folder. While the folder is on a desk, it can remind the user of the urgency of the documents within it. However, when the user puts the folder into a filing cabinet, the user can no longer see the urgency denoted by the red colour. A similar effect applies to digital documents. For example, saving an email attachment in a file folder removes the contextual information (e.g., from whom and when the document was received), which is usually available in the email body [20].

Context can also be lost because of some habitual use of computer applications. For example, an email’s subject and its content are not necessarily consistent, as users often simply press “reply” to an old message when starting new conversational threads. In a long conversational thread, users may not include all previous messages when sending a reply. They may also discuss topics that are irrelevant to the initial topic of discussion without changing the email subject.

3.3 Information Overload

People often have to deal with amounts of information exceeding their processing capability. This phenomenon appears in various studies of management of files [9], paper archives [21], email [14], and social networks [22, 23]. Across these studies, the main problem is related to the volume of information that people have to manage. On a closer look, however, there are finer issues in the information overload problem as discussed below.

People need to manage many ephemeral documents [9], such as emails and memos. Ephemeral documents usually serve as reminders, so they should be kept visible. However, computer screens are spatially limited, so users cannot arrange such documents freely. Moreover, the lifetime of these documents sometimes depends on other people. For example, a note is only relevant until a reply from another person arrives. Unfortunately, not all people respond to others' requests timely, making the management of ephemeral documents more difficult.

Studies [14, 24] observe that people use email not only for communication, but also for reminders, personal archiving, and task and contact management. These practices exacerbate the information overload problem, as they cause users to receive and manage more messages without using the right tool.

Another problem related to information overload, called premature filing, is observed in the management of paper archives [21]. This problem is described as follows. When receiving many documents, people cannot always decide the usefulness of these documents immediately. While they do not want to keep useless documents, they consider the potential value of these documents and hence hesitate to discard them right away. So they decide to file these documents, hoping to find time to assess their value later. Unfortunately, filing makes documents less visible. As people receive many new documents, they often forget about their filed documents. As a result, they keep many documents, which have little value to them, and are often discarded later without ever being read [21].

3.4 Interoperability

Problems in personal information management escalate when people have to share documents. People are familiar with their personal workspace: they organize it personally, know the contents of their document collections well, and decide the semantic organization of their workspace by themselves. These characteristics, however, do not exist in a shared repository: several people have access to the repository; others may add new documents to it; and as a team, they have to agree to some convention of how to organize this repository.

Due to the rich semantics in natural languages, people often use different words to refer to the same concept, and use the same word to refer to different concepts. According to Furnas et al. [25], the probability of two persons choosing the same word to describe the same concept is less than 0.20. In other words, it is unlikely that people will use the same vocabulary to describe the same things.

In collaborative work involving different groups of people, each group may require local customizations of documents, have different views of the shared objects, or favour a certain filing scheme [26, 27]. Dourish et al. [27] report that in an organization that has a standard filing scheme, each group still needs to adjust the standard filing scheme to meet requirements of different projects. As each group has its own local view, document sharing between groups becomes complicated. To be understandable, local customizations should be presented to each group differently, ideally, according to the corresponding group's local view.

Even for a small, relatively homogeneous team [28], interoperability problems exist. These problems include personal preferences over coarse- or finegrained categorization, topic- or purpose-based categorization, and syntactic- or semantic-based categorization.

4 Approaches to Handling Problems in Information Management

A variety of approaches have been proposed to deal with the problems discussed in the previous section. We focus on four main approaches below: tagging, flexible collections, temporal-based organizations, and the Semantic Desktop.

4.1 Tagging

The basic principle of tagging is to assign attributes (i.e., ⟨field, value⟩ pairs) to documents, and then to allow users to use these attributes to organize and retrieve documents [29–31]. This principle enables users to assign multiple attributes to a document, so they can express the semantics of a document more flexibly. Users are no longer limited to having to classify a document into a single category as with filing.

The concept of tagging has been applied to several file systems [29, 30, 32]. These systems use tagging for providing a flexible way of retrieving and grouping files. To retrieve and group files, users are no longer restricted to one strict hierarchical structure. Instead, they can group files based on a certain attribute of the files. For example, in a Semantic File System [29], to retrieve all papers about Semantic Web, a user can submit a query "topic=semantic web"[1] and the system will create a virtual directory containing all files that match the query.

Attributes have been used not only for providing associative access in file systems, but also for providing personalized services [33] and managing documents [31, 34, 35]. In general, these systems use attributes to capture metadata and to provide a flexible way of organizing and retrieving documents. Placeless Document Systems [34], however, extend this usage by allowing an attribute's value to be active code. So, for example, users can assign a program to a document, which backups the document periodically. This allows to associate not just semantics, but also computation with the corresponding documents.

4.2 Flexible Collections

Collections are an abstraction for handling the limitations of a hierarchical structure and for providing a flexible way of organizing various types of documents. There are two main principles of collections [31, 36]: (1) a document can belong to multiple collections, and (2) a collection can contain different types of documents.

The first principle allows a document to become a member of several different collections. This membership can be static, dynamic, or a combination of static

[1] The actual form of queries in Semantic File Systems is specified as virtual directory names, e.g., /sfs/topic:/semanticweb.

and dynamic. For example, semantic directories [30] and fluid collections [31] enable users to specify an inclusion list, an exclusion list, and a query to create a collection. The inclusion list contains a list of documents that must be included in the collection regardless of the query. The exclusion list specifies documents that must be excluded from the collection even if they match the query. The query states the criteria of documents to be included in the collection.

The second principle enables users to put various types of documents—such as emails, Web bookmarks, appointments, contact lists, and Word documents—together into a collection. Users can organize their documents in a more logical and meaningful way, as document organization is abstracted from the applications that produce the documents.

4.3 Time-Based Approaches

Time-based approaches provide an alternative to the desktop metaphor for managing information. These approaches are based on the fact that time is an important retrieval cue in information management [8, 17, 37]. Examples of such approaches include Lifestreams [38] and Time-Machine Computing [39].

Lifestreams [38] stores a user's personal information chronologically as a *lifestream*. A lifestream contains old documents, working documents, and possibly, future documents, such as appointments and reminders. A document's name is not mandatory because the system is responsible for identifying and placing the document in the lifestream. To facilitate information organization and retrieval, the user can create substreams to display only subsets of documents in the lifestream. Within a substream, the user can refine the query further by creating other substreams. Finally, Lifestreams is also capable of summarizing information in a substream and presenting the result to the user.

Time-Machine Computing (TMC) [39] extends the idea of Lifestreams [38] by capturing both temporal and spatial information of a document. To do so, TMC provides a special desktop that allows users to keep and organize their documents spatially on this desktop. Users may remove documents from their desktop. However, these documents are not deleted permanently because TMC keeps track of any state changes on the desktop. Thus, users can access their removed document by setting their desktop's time back to a point before they remove the desired document, and TMC will restore the state of the users' desktop at the specified time.

4.4 The Semantic Desktop

The core technology of the Semantic Desktop is the Semantic Web [4]: an extension of the current World Wide Web that enables machines to "understand" and process information intelligently. As discussed in [4], the main components of the Semantic Web consist of the Extensible Markup Language (XML), the Resource Description Framework (RDF), and ontologies. XML enables users to create arbitrary tags for describing the structure of documents. RDF provides a framework to express the meaning of information using XML. And an ontology—"an explicit specification of conceptualization" [40]—makes information sharing meaningful by providing a set of vocabulary to discuss a particular domain.

The main idea of the Semantic Desktop [3, 5, 6] is to apply principles of the Semantic Web [4] to personal information management. This approach allows the creation of semantically rich PIM tools. Semantic Desktop systems can use RDF to express both the structure and semantics of webs of information on a user's personal computer as metadata. Because this metadata can be "understood" by machines, this approach has potential for improving current practices in PIM: information overload can be reduced by delegating well-defined tasks to agents; information sharing can be more meaningful and contextually rich, as the corresponding metadata is also shared and described using common ontologies.

Like the Semantic Web, the Semantic Desktop treats information as a Web resource, so information is identified using a Uniform Resource Identifier [3, 6]. Some projects developing Semantic Desktop systems include Haystack [5], Gnowsis [6], Chandler,[2] and Fenfire.[3]

5 Design Criteria for Semantic Desktop Systems

To be adopted widely by users, Semantic Desktop systems must be designed properly. One of the most important aspects in design is to know the users. Thus, based on the existing user practices and limitations of approaches to personal information management, we propose preliminary design criteria for evaluating Semantic Desktop systems as follows:

- **Flexibility:** A Semantic Desktop system should support flexible information organization. This organization should be abstracted from both the structure of information storage (e.g., hierarchical file systems) and desktop applications (e.g., email tools, Web browsers, and task management tools). Grouping and regrouping information should be easy. Users should be able to arrange information based on their logical views, intentions, or usage patterns. For example, users should have options to sort information based on its frequency of access, its level of importance, or its level of urgency. Flexibility is important because users have different needs and preferences in organizing information (see section 2.2).
- **Retrievability:** Users should be able to find desired information easily. Just like other Web resources, users' personal information should be accessible virtually from anywhere. Users do not need to know where their data actually resides on. And since users like browsing their personal workspace [9, 11, 12], a Semantic Desktop system should provide a good browser, which supports multiple information visualizations and allows users to see cues and contextual information of their documents to facilitate retrieval. In other words, the search mechanism should favour recognition over recall.

[2] http://www.osafoundation.org/Chandler_Compelling_Vision.htm
[3] http://fenfire.org/vision.html

- **Security:** While easy access to information is desirable, security must not be compromised. Every access to the user's data must be authorized. Sensitive information should be transferred through a secure channel and encrypted. A Semantic Desktop system should authenticate shared data and metadata. Also, it should allow users to control access privileges of their information (e.g., public, group, or private data).
- **Context maintenance:** Contextual information of documents is usually stored as metadata. Creating metadata sometimes requires manual input from users. Unfortunately, users are generally reluctant to do more work, such as typing additional information about a document. Thus, a Semantic Desktop system should be able to capture or create metadata automatically, for example, by using context analysis [41] or data-mining techniques. If user input is unavoidable, then the system should provide a mechanism that allows easy annotation (e.g., [42]). Presenting summaries of interactions [43] can also help users recall the context of interactions while communicating with other people.
- **Proactiveness:** A Semantic Desktop system should be proactive in promoting awareness to the user. Examples include notifying users when they receive emails from their important contacts, highlighting important or related documents, and reminding users of their appointments and tasks.
- **Cognitive load:** A Semantic Desktop system should minimize the user's cognitive effort necessary to use the system. Reducing cognitive load can be achieved using various ways, such as encapsulating technical details from users, enabling users to delegate well-defined tasks to agents, presenting summaries of information, and filtering in/out incoming information.
- **Interoperability:** Interoperability can be examined from different perspectives. At a system level, a Semantic Desktop system should promote interoperability both among desktop applications on a single computer and with other Semantic Desktop systems. At a user level, the system should resolve inconsistencies among different ontologies used by users.
- **Performance:** A system performance should be reasonable and acceptable to users. Responses to user queries should be timely. Resource consumption and allocation should be well managed.

Evaluating a PIM tool rigorously requires a well-designed user study and in-depth analysis. While such rigorous analysis is out of scope of this paper, we present simple analyses to illustrate trade-offs of existing approaches to PIM.

Most operating systems and desktop applications currently offer hierarchical structures for information organization. Such structures are not *flexible* enough to meet the user requirements (see section 3.1). Consequently, users feel that organizing information is a heavyweight *cognitive* activity. This inflexibility also complicates information *retrieval*, as the *context* of documents can only be maintained in a limited way. Despite their disadvantages, hierarchical tree structures are computationally efficient and hence deliver good *performance*.

The Semantic Desktop views information, regardless of its type, as a Web resource. This view allows *flexible* information organization and easy *retrieval*.

This approach promotes a *contextually* rich PIM environment where metadata is described using RDF and ontologies, which support *interoperability*[4] among applications and with other systems. *Proactive* agents can be used to deal with information overload. The Semantic Desktop, however, is complex, and therefore should minimize the user's *cognitive* load so that it is usable by average users.

6 Social Aspects of Knowledge Communities

We have reviewed practices, problems, and approaches in information management from individual users' point of view. While this review is useful for developing a PIM tool, to build successful knowledge communities, we should also consider social aspects of communities. What is the role of social networks in information and knowledge management? How can we use social networks to develop a better infrastructure for knowledge management? How can users be motivated to participate and contribute to their communities? These issues are discussed in this section.

6.1 The Role of Social Networks

Social networks [44] study relationships and information flows among people, and play an important role in information and knowledge management. People set a high priority to messages that they receive from families, bosses, or closed friends. They prefer to collaborate with their trusted friends. And they are more willing to share their knowledge in an environment where they feel accepted and secured—safe to speak up their minds, safe to make mistakes, and so on. In other words, trust and interpersonal relationships are important. It is not surprising that people make considerable efforts in maintaining and expanding their social networks [22].

Informal conversations, which usually happen in office lounges or other recreational places, have potential for building trust and relationships [45]. In such relaxing places, people can talk informally about work or personal things while developing relationships with other co-workers. In computer systems, such informal conversations can be supported by chat applications [45, 46].

The main activities in maintaining social networks include remembering and communicating [22]. Remembering people's expertise is necessary for locating the right resources. Remembering personal details enables people to give a personal touch to their relationships, for example, by sending birthday or anniversary cards. To facilitate remembering, people maintain intermittent communications, especially with members of their social networks who are currently inactive, such as with collaborators from past projects. Communication is helpful to refresh memories of both parties and to maintain connections between them.

[4] Achieving large scale interoperability, however, is a non-trivial task.

6.2 User Motivation

Communities need a critical mass of users. Not only the number of users is crucial, but also their participation in the communities. Without a critical mass, the systems underlying communities will be abandoned by the users [47, 48].

Every person has a personal goal and motivation to join and participate in a knowledge sharing community. Some may want to meet new people with similar interests while others want to learn and follow the development of a research area. Although individual goals may vary, members of a community usually share a common interest to some degree. In general, people will stay active in a community if by doing so they can achieve their goals. Despite the variety of individual goals and motivation, how can we motivate users to participate actively in their communities?

Giving rewards to active users is a common approach to increasing user participation in online communities. Rewards can be in various forms, including getting better services, more privileges, or a higher status or visibility in the community [48].

Because one's behaviour is usually influenced by others, another approach to motivating users is by promoting social awareness within a community [47, 49]. The basic idea is to give information about the structure and activities in a community. Examples include displaying a list of online users, giving notifications about some events, and visualizing the users' reputation and social networks. By knowing that there are activities in their community, users are expected to stay active in it (i.e., to keep using the system). Furthermore, knowing the "presence" of others can give a good feeling to users [46]. Social awareness can also serve as social control [50] because when users realize that others know about their activities or behaviour, they will likely behave according to social norms.

6.3 Communities of Practice

In practice, the most effective ways of creating and communicating knowledge are not through written documents, but through personal interactions, including informal conversation, storytelling, and dialogue [45, 51]. Personal interactions are important because knowledge is not an isolated entity: knowledge is a concrete form of what is in a person's mind, and how others perceive it [51]. Thus, knowledge exchange processes involve both knowledge creators and users. When reading a document, readers can only rely on their perception. They cannot validate whether they have understood the document correctly. Interactions with the author can reduce this problem, as readers can argue, ask questions, or validate their understanding with the author. At the same time, the author can assess readers' interest and understanding by looking at their gestures or asking for feedback.

Because of the nature of knowledge exchange processes, people use knowledge databases not only to find specific information, but also to interact with people who produce or use the information [50]. For example, to increase the chance of getting a project proposal approved, it is very helpful to read examples of successful proposals and to talk with the authors. The authors can give many insights that are not available in the proposals, or help introduce key persons as potential referees. Thus, in addition to storing information, knowledge databases should help users maintain and expand their social networks [49].

7 Concluding Remarks

To design a collaboration infrastructure, we should consider both its technical and social aspects. Technologies are a prerequisite for implementing such infrastructures. However, whether users will adopt and use an infrastructure often depends on design subtleties, such as how well the design fits the existing user practices. While we cannot control the dynamics of communities, we can support them by designing an infrastructure that fosters characteristics of successful communities.

In this paper, we focused mainly on personal information management and proposed design criteria for Semantic Desktop systems. However, we also brought social issues in knowledge management to the discussion, as these factors are crucial in the success of communities.

Acknowledgments. We thank Joyce Boedianto, Jeff Smith, and Yao Wang for giving comments and suggestions on drafts of this paper. The development of design criteria for Semantic Desktop systems is based on a suggestion from an anonymous reviewer.

References

1. Kollock, P.: Design principles for online communities. In: Harvard conf. on the Internet and society. (1996)
2. Davies, J.: Supporting virtual communities of practice. Industrial Knowledge Management (2001)
3. Decker, S., Frank, M.: The networked semantic desktop. In: Workshop on application design, development and implementation issues in the semantic web. (2004)
4. Berners-Lee, T., Hendler, J., Lassila, O.: The semantic web. Scientific American (2001)
5. Huynh, D., Karger, D.R., Quan, D.: Haystack: a platform for creating, organizing and visualizing information using RDF. In: Semantic Web Workshop. (2002)
6. Sauermann, L.: The gnowsis semantic desktop for information integration. In: 1st Workshop on Intelligent office appliances. (2005)
7. Schmitz, C.: Towards self-organizing communities in peer-to-peer knowledge management. In: Workshop on Ontologies in peer-to-peer communities. (2005)
8. Malone, T.W.: How do people organize their desks? Implications for the design of office information systems. ACM Trans. Inf. Syst. **1** (1983) 99–112
9. Barreau, D., Nardi, B.A.: Finding and reminding: file organization from the desktop. SIGCHI Bull. **27** (1995) 39–43
10. Mander, R., Salomon, G., Wong, Y.Y.: A 'pile' metaphor for supporting casual organization of information. In: Proc. conf. on Hum. factors in comp. sys. (1992) 627–634

11. Ravasio, P., Schär, S.G., Krueger, H.: In pursuit of desktop evolution: user problems and practices with modern desktop systems. ACM Trans. Comput.-Hum. Interact. **11** (2004) 156–180
12. Boardman, R., Sasse, M.A.: "Stuff goes into the computer and doesn't come out": a cross-tool study of personal information management. In: Proc. conf. on Hum. factors in comp. sys. (2004) 583–590
13. Abrams, D., Baecker, R., Chignell, M.: Information archiving with bookmarks: personal web space construction and organization. In: Proc. conf. on Hum. factors in comp. sys. (1998) 41–48
14. Whittaker, S., Sidner, C.: Email overload: exploring personal information management of email. In: Proc. conf. on Hum. factors in comp. sys. (1996) 276–283
15. Kidd, A.: The marks are on the knowledge worker. In: Proc. conf. on Hum. factors in comp. sys. (1994) 186–191
16. Kwasnik, B.: How a personal document's intended use or purpose affects its classification in an office. In: Proc. 12th ACM SIGIR conf. on Research and development in information retrieval. (1989) 207–210
17. Rodden, K., Wood, K.R.: How do people manage their digital photographs? In: Proc. conf. on Hum. factors in comp. sys. (2003) 409–416
18. Lansdale, M.: The psychology of personal information management. Applied Ergonomics **19** (1988) 55–66
19. Chase, W.G., Simon, H.A.: Perception in chess. Cognitive Psychology (1973) 55–81
20. Bondarenko, O., Janssen, R.: Documents at hand: learning from paper to improve digital technologies. In: Proc. conf. on Hum. factors in comp. sys. (2005) 121–130
21. Whittaker, S., Hirschberg, J.: The character, value, and management of personal paper archives. ACM Trans. Comput.-Hum. Interact. **8** (2001) 150–170
22. Nardi, B.A., Whittaker, S., Schwarz, H.: Networkers and their activity in intensional networks. Computer Supported Cooperative Work **11** (2002) 205–242
23. Whittaker, S., Jones, Q., Nardi, B., Creech, M., Terveen, L., Isaacs, E., Hainsworth, J.: ContactMap: organizing communication in a social desktop. ACM Trans. Comput.-Hum. Interact. **11** (2004) 445–471
24. Bellotti, V., Ducheneaut, N., Howard, M., Smith, I.: Taking email to task: the design and evaluation of a task management centered email tool. In: Proc. conf. on Hum. factors in comp. sys. (2003) 345–352
25. Furnas, G.W., Landauer, T.K., Gomez, L.M., Dumais, S.T.: The vocabulary problem in human-system communication. Commun. ACM **30** (1987) 964–971
26. Simone, C., Mark, G., Giubbilei, D.: Interoperability as a means of articulation work. In: Proc. conf. on Work activities coord. and collaboration. (1999) 39–48
27. Dourish, P., Lamping, J., Rodden, T.: Building bridges: customisation and mutual intelligibility in shared category management. In: Proc. intl. ACM SIGGROUP conf. on Supporting group work. (1999) 11–20
28. Berlin, L.M., Jeffries, R., O'Day, V.L., Paepcke, A., Wharton, C.: Where did you put it? Issues in the design and use of a group memory. In: Proc. conf. on Hum. factors in comp. sys. (1993) 23–30
29. Gifford, D.K., Jouvelot, P., Sheldon, M.A., O'Toole, J.W.: Semantic file systems. In: Proc. 13th ACM symp. on Operating systems principles. (1991) 16–25
30. Gopal, B., Manber, U.: Integrating content-based access mechanisms with hierarchical file systems. In: Proc. 3rd symp. on Operating systems design and implementation. (1999) 265–278
31. Dourish, P., Edwards, W.K., LaMarca, A., Salisbury, M.: Presto: an experimental architecture for fluid interactive document spaces. ACM Trans. Comput.-Hum. Interact. **6** (1999) 133–161

32. Bowman, C.M., Dharap, C., Baruah, M., Camargo, B., Potti, S.: A file system for information management. In: Proc. ISMM intl. conf. on Intelligent information management systems. (1994)
33. Adar, E., Kargar, D., Stein, L.A.: Haystack: per-user information environments. In: Proc. 8th intl. conf. on Information and knowledge management. (1999) 413–422
34. Dourish, P., Edwards, W.K., LaMarca, A., Lamping, J., Petersen, K., Salisbury, M., Terry, D.B., Thornton, J.: Extending document management systems with user-specific active properties. ACM Trans. Inf. Syst. **18** (2000) 140–170
35. Hopkins, I., Vassileva, J.: Beyond keywords and hierarchies. Journal of Digital Information Management **3** (2005) 139–145
36. Karger, D.R., Quan, D.: Collections: flexible, essential tools for information management. In: CHI '04 extended abstracts. (2004) 1159–1162
37. Dumais, S., Cutrell, E., Cadiz, J., Jancke, G., Sarin, R., Robbins, D.C.: Stuff I've seen: a system for personal information retrieval and re-use. In: Proc. 26th ACM SIGIR conf. on Research and development in information retrieval. (2003) 72–79
38. Freeman, E., Gelernter, D.: Lifestreams: a storage model for personal data. SIGMOD Rec. **25** (1996) 80–86
39. Rekimoto, J.: Time-machine computing: a time-centric approach for the information environment. In: Proc. 12th annual ACM symp. on User interface software and technology. (1999) 45–54
40. Gruber, T.R.: Toward principles for the design of ontologies used for knowledge sharing. Int. J. Hum.-Comput. Stud. **43** (1995) 907–928
41. Soules, C.A.N., Ganger, G.R.: Why can't I find my files? New methods for automating attribute assignment. In: Proc. 9th Workshop on Hot topics in operating systems. (2003)
42. Gemmell, J., Bell, G., Lueder, R., Drucker, S., Wong, C.: MyLifeBits: fulfilling the memex vision. In: Proc. 10th ACM intl. conf. on Multimedia, New York, NY, USA, ACM Press (2002) 235–238
43. Fisher, D., Dourish, P.: Social and temporal structures in everyday collaboration. In: Proc. conf. on Hum. factors in comp. sys. (2004) 551–558
44. Garton, L., Haythornthwaite, C., Wellman, B.: Studying online social networks. Doing Internet Research (1999) 75–105
45. Thomas, J.C., Kellogg, W.A., Erickson, T.: The knowledge management puzzle: human and social factors in knowledge management. IBM Systems Journal **40** (2001) 863–884
46. Nardi, B.A., Whittaker, S., Bradner, E.: Interaction and outeraction: instant messaging in action. In: Proc. ACM conf. on CSCW. (2000) 79–88
47. Ackerman, M.S., Starr, B.: Social activity indicators: interface components for cscw systems. In: Proc. 8th annual ACM symp. on User interface and software technology. (1995) 159–168
48. Cheng, R., Vassileva, J.: User motivation and persuasion strategy for peer-to-peer communities. In: Proc. 38th Hawaii intl. conf. on System sciences. (2005)
49. Bush, A.A., Tiwana, A.: Designing sticky knowledge networks. Commun. ACM **48** (2005) 66–71
50. Erickson, T., Kellogg, W.A.: Social translucence: an approach to designing systems that support social processes. ACM Trans. Comput.-Hum. Interact. **7** (2000) 59–83
51. Walsham, G.: Knowledge management: The benefits and limitations of computer systems. European Management Journal **19** (2001) 599–608