No More Isolated Files: Managing Files as Social Artifacts

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

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

CHI 2006, April 22–27, 2006, Montréal, Québec, Canada.
ACM 1-59593-298-4/06/0004.

Abstract
People manage and use their files/documents in social settings. However, current file systems do not consider social interactions that happen around files. For example, files that have been sent to many people are displayed in the same way as those that have never been sent. We argue that by capturing user activities around files, we can enhance user experience in managing their documents. In particular, presenting activity traces can help users comprehend their document collections better and recall the context of their documents. In this paper, we discuss our approach to providing a context-rich environment for document management.

Keywords
Document management, personal information management, semantic desktop, user modelling, context, metadata

ACM Classification Keywords
H.3.2 [Information Storage and Retrieval]: Information Storage—file organization; H.4.3 [Information Systems Applications]: Communication Applications—information browsers; H.5.2 [Information Interfaces and Presentation]: User Interfaces—graphical user interfaces

Introduction

People manage and use their personal information—documents, email messages, web bookmarks, to-do lists—in rich social settings. They write papers for specific conferences, create slides to present their ideas to certain audience, and collaborate with their colleagues to complete some tasks. The way people handle information is shaped by such contexts. For example, people may use a red folder to keep urgent documents while using a green folder to keep other documents.

Contextual information about a document can refer to various things. Among others, it may refer to the purpose, the intended readers, or the status of the document. It describes the circumstances that help in understanding a document.

Context plays an important role in personal information management (PIM). Unfortunately, current file systems treat files as isolated entities. Consider an email attachment. The email message usually contains contextual information about the attached document: the sender, the subject, the date, the status (e.g., draft, final version), and so on. However, when the user saves this attachment in a file folder, the contextual information is lost, and the attachment becomes an isolated entity in the file system. Thus, to avoid this problem, users prefer to keep many email messages in their mailboxes [2]. This practice contributes to the email overload problem [11], where users extend the use of email from a communication tool to personal archiving, reminders, and task management.

We aim to improve current file systems by developing a tool that monitors user activities at the computer, collects relevant information as metadata, and presents it appropriately to users to help them recall contextual information about their files. In this paper, we discuss problems in current file systems and our proposed solution, including its potential implications. We also discuss potential challenges in implementing our approach and directions for future work.

Problem Definition

Context is essential in information management; it plays an important role in the process of recall [7]. People use various cues—physical, spatial, temporal, social—to maintain contextual information about their documents. The role of spatial cues, for example, has been observed in office organization [2, 8], where people keep recent or urgent documents closer to the centre of their working area. Such spatial organization is also adopted in digital information management. Computer users arrange their computer desktop systematically [9]: they group program shortcuts in a specific region and put related documents close to one another. Such arrangements can help users locate frequently used programs or documents on their computer.

Unfortunately, current file systems do not offer a context-rich environment for PIM. For example, by looking solely at the file, as displayed in a file browser, one cannot tell whether a project report has been circulated to the project members. If file systems trace such user activities (e.g., sending a file to a group of people), then the systems can present this information to users so that they can comprehend their document collections better. To cope with this problem, users

currently use a pragmatic, limited way to express contextual information about a file by giving it a meaningful name and saving it in an appropriate folder.

Thus, the main problem that we want to address here is the lack of contextual information about files in file systems. In particular, we aim to trace user activities around files and use this information to enhance current file systems. Our hypothesis is that by capturing and presenting metadata of files, we can improve the user's experience in managing documents.

Proposed Solution

Our goal is to develop a tool that can enhance PIM. In particular, the tool is aimed to monitor user activities, collect and capture relevant information as metadata, and present the metadata appropriately to help users comprehend and manage their documents. Figure 1 illustrates examples of activities that can be monitored to reveal meaningful information to the user. These use cases lead to the following requirements.

Requirements

The first question in pursuing our goal is, hence, what kinds of metainformation about files that can be useful for users. Current file systems maintain basic metadata of files, such as type, size, author, date created, last modified, and last accessed. Such basic metadata is insufficient to capture the richness of social interactions that happen around files. Furthermore, although users have options to annotate their files, file browsers only display basic metadata, which cannot reveal user activities around the files. Therefore, more metadata is necessary for improving current file systems, including the following:

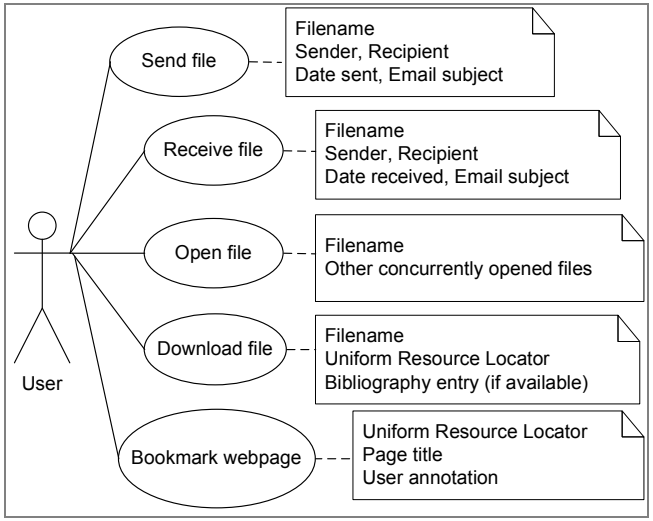

figure 1. Examples of user activities and their potentially useful metadata.

- Social information. Because files are used in social settings, collecting social information about files can be useful. Tracking user activities such as sending or receiving a file to or from another person can reveal meaningful information to the user. Such information may facilitate users to recall work context of their files and to connect with other people.
- Temporal information. While working on a task, people often need to access information from different sources. Let's consider writing a conference paper. During the writing process, people may work using their word processor while consulting their bibliography file or a document that describes formatting requirements for the conference. There are multiple, related documents that are opened concurrently. Such access patterns can then be used by file browsers to

highlight related documents, which can serve as reminders or retrieval cues for users. Other temporal information that may be useful includes recently accessed files and sequential access pattern—files that are accessed in sequence one after the other.

- Other information. Besides social and temporal information, other potentially useful metadata of files includes the original locations of downloaded files (i.e., Uniform Resource Locators), user annotations, frequently accessed files, bibliography entries of research papers, and percentage of changes made to files. Some of this information can be acquired automatically, which reduces user effort in managing documents.

Design of the system

Our approach is to develop a tool, rather than a new file system, to enhance current file systems. Thus, our metadata repository is independent from the file systems. As illustrated in figure 2, our system design consists of three main components: observer, metadata repository, and presentation.

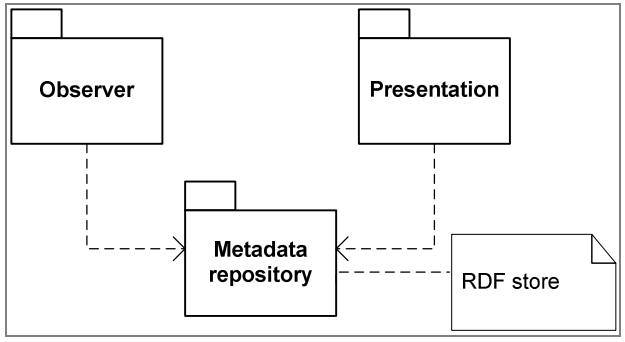

figure 2. Design of the system.

- The main function of the observer component is to monitor user activities. When certain events occur (e.g., sending a file), this component captures the events and puts some information (e.g., filename, recipient, date sent) in the metadata repository.
- The metadata repository uses the Resource Description Framework (RDF) [10] to represent metadata. We choose RDF for the following reasons. First, RDF is a framework recommended by the World Wide Web Consortium [12] for representing information in the Semantic Web [1]. Using a standard format like RDF promotes interoperability among different systems, as metadata can be exchanged seamlessly. Second, RDF is machine "understandable." Thus, mundane, well-defined tasks can be delegated to software agents to reduce users' workload.
- The presentation component retrieves relevant metadata from the repository and presents it to users. We plan to develop a file browser or integrate this component into an existing file browser to allow users to use a familiar tool while enjoying a more context-rich environment for managing their documents.

Related Work

Previous research [5] has shown that there are meaningful interaction patterns that can be traced from email usage. Some of these patterns are easily comprehensible to the user. Furthermore, these patterns are useful for helping users recall the context of their conversations with other people. This research, however, focuses only on extracting interaction patterns from email use.

Attempts to improve PIM also come from the semantic desktop community [4]. The basic principle is to apply

Semantic Web technologies (e.g., RDF, ontologies) to PIM [3, 6]. A well known example of semantic desktop systems is Haystack [6], which focuses on providing an integrated, customizable PIM tool.

Our approach takes inspiration from the research discussed above. However, our focus is different. Instead of focusing on a particular application (e.g., email), we attempt to extract useful patterns of user activities from various applications, including email, web browser, instant messenger, and word processor. The observer component, as discussed in the previous section, may also gather additional information (e.g., bibliography entry) proactively. Thus, our focus is to enrich PIM by capturing and presenting user-activity traces as metadata appropriately. To promote interoperability, we use RDF to represent the metadata.

Discussion

To conclude this paper, in this section we discuss potential implications and challenges of our approach, and our plan for future work.

Potential implications

Potential implications of our approach include the following:

- Better information understanding. People manage and use information in certain contexts. Presenting activity traces, such as sending or receiving a file, can help users recall the context of their documents. For example, by looking at a list of recipients of a file, the user may be able to recall a lot of information about the file immediately, such as what the file is about, what the status of the document is, and who is involved in a particular project. This information can also serve as retrieval cues, which aid the user to find desired information.
- Better information sharing. RDF allows the creation of links between resources (i.e., information). Thus, when it comes to information sharing, users can share not only their documents, but also the associated metadata [3]. This metadata can be user annotations or social information about the documents, contact information of the authors, or bibliography entries of research papers. Information sharing hence becomes more meaningful, as contextual information is also shared to the recipient. Furthermore, because this metadata is written using a standard format, instead of a proprietary format, integration with local systems or applications becomes easier. As a result, users do not have to recreate the already available metadata on their local computers; they can just import and reuse it.
- Easy integration with the Semantic Web. Most of today's personal computers have become an access point to the World Wide Web; they are no longer standalone devices. Being able to create links between pieces of information, regardless of its type or location, will enable users to create semantic networks. Information storage will become transparent. And computers will turn into a medium to express users' ideas or knowledge, not just an information-processing device.

Potential challenges

Observing user activities to discover useful patterns has its potential challenges, including the following:

- Detecting context switching. Monitoring temporal access patterns of documents may lead to linking completely irrelevant documents. While working on a

document, users sometimes access unrelated documents, for example, to take a break. This problem may be solved using content analysis of concurrently accessed documents to measure their similarity.

- Presenting metadata. Different files have different kinds of metadata. Designing a user interface that can present various types of metadata elegantly can be challenging. Moreover, while we want to present relevant information about files, we do not want to give too much information to users, especially information that is irrelevant to the current user work context, to avoid information overload.

Future work

This project is still in an early phase. We have defined the problem, gathered initial requirements, and proposed our design solution. As future work, we are going to refine and implement our design, and then conduct a user study to evaluate the system. The evaluation criteria include the following aspects: usability, usefulness of captured metadata, and system performance.

References

[1] Berners-Lee, T., Hendler, J., and Lassila, O. The semantic web. *Scientific American* (2001).

[2] Bondarenko, O. and Janssen, R. Documents at hand: Learning from paper to improve digital technologies. In *Proc. CHI 2005*, ACM Press (2005), 121-130.

[3] Decker, S. and Frank, M. The networked semantic desktop. In *Proc. WWW 2004 Workshop on Application Design, Development and Implementation Issues in the Semantic Web* (2004).

[4] Decker, S., Park, J., Quan, D., and Sauermann, L. (Eds.). *Proc. ISWC Workshop on the Semantic Desktop* (2005).

[5] Fisher, D. and Dourish, P. Social and temporal structures in everyday collaboration. In *Proc. CHI 2004*, ACM Press (2004), 551-558.

[6] Huynh, D., Karger, D., and Quan, D. Haystack: A platform for creating, organizing and visualizing information using RDF. In *Proc. WWW 2002 Semantic Web Workshop* (2002).

[7] Lansdale, M. The psychology of personal information management. *Applied Ergonomics 19*, 1 (1988), 55-66.

[8] Malone, T.W. How do people organize their desks? Implications for the design of office information systems. *ACM Transactions on Information Systems 1*, 1 (1983), 99-112.

[9] Ravasio, P., Schär, S.G., and Krueger, H. In pursuit of desktop evolution: User problems and practices with modern desktop systems. *ACM Transactions on Computer-Human Interaction 11*, 2 (2004), 156-180.

[10] Resource Description Framework (RDF). http://www.w3.org/RDF/

[11] Whittaker, S. and Sidner, C. Email overload: Exploring personal information management of email. In *Proc. CHI 1996*, ACM Press (1996), 276-283.

[12] World Wide Web Consortium. http://www.w3.org/.