Just Email It to Me! Why Things Get Lost in Shared File Repositories

Emilee J. Rader

University of Michigan, School of Information
1075 Beal Ave., Ann Arbor MI, 48109
ejrader@umich.edu

Dissertation Advisor:
Judith Olson [jsolson@umich.edu]

Abstract. Shared file repositories are a type of information technology application used by workgroups to store and share files online. Their use in organizations is becoming more frequent; however, repository users are not always able to effectively find and access information, especially when files in the repository have been created and maintained by others. Through field studies involving current users of shared file repositories, I will document and analyze the scope and consequences of the problem. In addition, I will test hypotheses about possible remedies through a series of experiments exploring the effects of common ground on folder hierarchy and naming structure, and the ability of users to find and access files.

Keywords. Shared file repositories, social computing, information management, common ground

1 Introduction

Much of an organization's information is represented in the form of documents, such as reports, memos, meeting minutes, email messages, etc. Ineffective document management incurs costs such as "lost work time, ineffective access to information, duplication of effort, failure to share information, and information overload" [6]. Many different kinds of workgroups including research labs, corporate teams, and software developers use central online repositories for storing information. These repositories are maintained by many organizations, "for their potential value in the day-to-day operations of the organization" [8]. They are essential for document sharing, and can be greatly beneficial for organizational efficiency, communicating organizational goals, and also for learning and innovation. They can contain "mission critical information" such that if it were lost there would be serious consequences [2].

Despite the importance of the information stored within them, shared file repositories generally do not have explicit rules or structures for organization and searching, like a library catalog does. Instead, they tend to accumulate content over time and become more and more disorganized, such that users have difficulty finding the files they need. The research described below examines situations in which users are not able to effectively find and access information in shared file repositories, and suggests possible remedies for these problems.

2 Shared File Repositories

Shared file repositories are online storage spaces used by workgroups for storing, organizing, and sharing documents and other files, and their use is increasing. A shared file repository is more complex than just an "aggregate of every individual's contribution" [7], and maintaining it is a collaborative activity. This makes repositories different from other ways shared work files are commonly stored, such as in email mailboxes, on personal computers, or posted to internal company websites. People differ in the ways in which they structure their personal file repositories, as has been observed in many personal information management studies; for example, [1]. Individual goals and strategies for managing information could affect the choices users make when storing or seeking files in a shared file repository. A repository user is generally familiar with his fellow group members, and with projects and joint work activities they are engaged in together. However, he can expect to be familiar with only some of the files stored in a shared file repository, and he may or may not have been involved with creating the hierarchy and naming structure, or with storing and moving files around in the repository. This creates a situation different from both searching the web and one's personal information, where a user might be trying to find files with which she is unfamiliar, or looking for familiar files stored in unfamiliar places. This can be frustrating enough that users seeking information circumvent shared file repositories altogether, opting to request files from others via email instead – thus the title of my dissertation.

3 Common Ground

The functionality available in most shared file repositories is essentially identical to personal information management software, with additional capabilities allowing multiple people to interact with them simultaneously. However, the communication and collaboration inherent to group work lead to problems not accounted for in the design of personal information management tools, nor addressed in the personal information management literature. For example, previous research has demonstrated that if two random users were to create a name for the same file, it is unlikely that they would choose

identical words [4]. Fortunately, users of shared file repositories are not necessarily random pairs of people who are unknown to each other. In the best case, they share a work context and even have some knowledge about each other's preferences and personal styles. So, while there is naturally a great deal of variability in people's choices when storing files in a shared file repository, their knowledge about each other and their shared context – their common ground – might mitigate the problem somewhat, if it were somehow brought to bear.

Common ground [3] is the mutual knowledge, beliefs and assumptions that people share about each other. It is inferred based on joint membership in cultural communities and through shared perceptual experiences, and accumulates via conversation. As conversation progresses, people introduce ideas and vocabulary that become part of their common ground, and can subsequently be referred to without the overhead of having to re-introduce them. There is much experimental evidence to support the idea that common ground affects language use. Speakers tailor their utterances for listeners, with performance implications. In addition, people create labels for their own use that are different from those created for an unknown future person [5]. People tailor what they say to whomever is the intended recipient; it is reasonable to think that common ground might indeed affect the names information producers create for files they store in a shared file repository.

4 Proposed Research

In my dissertation, I build on the current understanding of personal information management tasks and tools by examining similar tasks in group situations having multiple participants and stakeholders. CTools (ctools.umich.edu) is an example of software used at the University of Michigan that includes shared file repository functionality. I conducted a pilot study with faculty, staff and students using CTools to support ongoing collaborative projects, in which I collected data through interviews and server event log analysis.

Through this continuing field study, I will obtain a more detailed understanding of the problems users encounter under real work conditions when using a shared file repository, including how decisions are made about what to share via a repository, how files should be named, and where they should be stored. I will also document the scope and consequences of the problems users have with finding and accessing files. The field study will include additional semi-structured interviews, content analysis of the structure, naming conventions, and information contained within shared file repositories, and further analysis of usage log data providing information about user behavior and structural changes to the repositories over time. In addition, I will test hypotheses about possible remedies through a series of experiments exploring the effects of common ground, made salient by various feedback and incentive mechanisms, on choices made by users when storing, organizing, and seeking files using shared file repositories.

I am also studying the properties and usage of a different kind of online application for sharing information, the popular social bookmarking and tagging website, del.icio.us (http://del.icio.us). This system provides the capability for users to bookmark web pages and associate user-generated metadata, or tags, with them. Through this project I am able to study an application used by millions of people, and gain valuable insight into design alternatives to the traditional hierarchical file-and-folder structure for storing information.

Shared file repositories are typically not considered to be communications technologies in the same way email and instant messaging are. Yet these applications support asynchronous interactions and information exchanges between workgroup members. Through choices about which files are stored in close proximity, who is given access, and via metadata such as file author and usage frequency, repositories can communicate information to others about which activities are more important, how projects are related, and even the power relationships within the workgroup. This research will provide a better understanding of social interactions mediated by shared file repositories in workgroup settings, as well as suggesting technological and social interventions that can be used to inform the design and development of future applications.

References

1. Berlin, L.M., Jeffries, R., O'Day, V.L., Paepcke, A. and Wharton, C.: Where did you put it? Review of Information Science and Technology, Vol. 37. The American Society for Information Science and Technology, Medford, NJ (2002) 3-50
2. Blair, D.C. and Kimbrough, S.O.: Exemplary documents: a foundation for information retrieval design. Information Processing and Management, 38:363-379, (2002)
3. Clark, H.H.: Common Ground. Using Language. Cambridge University Press, Cambridge (1996)
4. Furnas, G.W., Landauer, T.K., Gomez, L.M. and Dumais, S.T.: The vocabulary problem in human-system communication. Commun. ACM, 30(11):964-971, (1987)
5. Fussell, S.R. and Krauss, R.M.: The effects of intended audience on message production and comprehension: Reference in a common ground framework. Journal of Experimental Social Psychology, 25(3):203-219, (1989)
6. Gordon, M.D.: It's 10 a.m. do you know where your documents are? The nature and scope of information retrieval problems in business. Information Processing & Management, 33(1):107-122, (1997)
7. Jian, G. and Jeffres, L.: Understanding Employees' Willingness to Contribute to Shared Electronic Databases: A Three Dimensional Framework. Communication Research, 33(4):242-261, (2006)
8. Trigg, R.H., Blomberg, J. and Suchman, L.: Moving document collections online: The evolution of a shared repository. In: Kyng, M., Bodker, S.andSchmidt, K. (eds.): *In Proc. Sixth European Conference on Computer-Supported Cooperative Work*. pp. 331-350. Kluwer Academic Publishers, Copenhagen, Denmark (1999)