

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

Emilee Rader
School of Information
University of Michigan
ejrader@umich.edu

ABSTRACT

Shared file repositories are a type of user-contributed content 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 in group information management systems.

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" [7]. Many different kinds of workgroups including research labs, corporate teams, and software developers use central online repositories for storing information. Some examples include enterprise content management systems, software code repositories, and wikis or blogs used by teams as a form of "group memory."

Online repositories are maintained by many organizations. 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 remedies for these problems.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

GROUP '07, November 4-7, 2007, Sanibel Island, Florida, USA
Copyright 2007 ACM 1-58113-000-0/00/0004 ...\$5.00

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" [8], and maintaining it is a collaborative activity. 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 sometimes to request files from others via email instead.

Several aspects of shared file repositories set them apart from other ways information is stored and shared online. Similar to personal information management, users fill both information producer and consumer roles. However, in a shared file repository producers and consumers are not necessarily the same person. Because people differ in the ways in which they structure their personal file repositories, this can be a problem for information seeking [1, 11]. In addition, shared file repositories generally do not have a structured organization scheme, unlike libraries. Library records and classification schemes were created for describing items and codifying relationships between subjects [10]. In contrast, shared file repositories do not necessarily have unified goals or purposes to guide users' choices. Being *unstructured* means that less effort is required when storing and labeling files.

However, the lack of structure can also have a down side. The act of storing and labeling files in a shared file repository is one of packaging content for later reuse, and in most situations, people do not effectively package content for others [9]. Effective packaging requires that one consider the information needs and context of the reusers, or information consumers. Information producers' unstructured choices can affect the future reuse of the information in the repository by information consumers, because decisions about file names and locations provide the means by which the user-contributed content may be found and accessed. The distinction between those producing the content and those consuming it is an important one: it implies that there is communication going on between users who are producing content, and users who are accessing content, via the repository. And because file and

folder hierarchies are represented in the form of text, factors that shape language choices in communication situations might also affect packaging choices for user-contributed content.

3. LANGUAGE USE

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 [5]. 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 [6]. 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. An important difference between a real-time conversation and any type of asynchronous communication is in the timing of feedback, which is important for establishing common ground and negotiating meaning [4]. For example, facial expressions are a form of nonverbal feedback that convey whether or not a speaker has been understood by a listener. Systems allowing for the provision of user-contributed content rarely include a mechanism for feedback; indeed, what information might be included in that feedback, and what form it might take, are open questions.

4. PROPOSED RESEARCH

The purpose of this research is to improve users' ability to find information in situations where multiple users access shared information resources via a user-contributed content system, by investigating how language use affects the organization, structure, and seeking of information. In my dissertation, I will answer the following questions:

1. How do groups organize and share information via an online repository of user-contributed content? What problems do individual members encounter during information seeking within the repository?
2. How do the source of common ground and the intended audience affect file and folder name choices, and the effectiveness of those names for information seeking?

I will first conduct a field study a user-contributed content system at the University of Michigan (ctools.umich.edu). I completed pilot

interviews with faculty, staff and students using CTools to support ongoing collaborative projects. I will follow up the pilot with a more focused investigation of information seeking by group members within their own shared file repositories. I am also planning a series of experiments subsequent to the field study that will test hypotheses about the effects of common ground on choices made by users when storing, organizing, and seeking files using shared file repositories.

This dissertation will add to our understanding of language use in situations not traditionally thought of as communication. I will explore factors that affect the usage and usefulness of a growing category of CSCW systems (user-contributed content), by combining psychology and information science perspectives to investigate unstructured organization schemes. Results will be used to inform system architecture and interface design choices.

5. REFERENCES

- [1] L. M. Berlin, R. Jeffries, V. L. O'Day, A. Paepcke, and C. Wharton. Where did you put it? issues in the design and use of a group memory. In *SIGCHI conference on Human factors in computing systems*, Amsterdam, The Netherlands, 1993.
- [2] D. C. Blair and S. O. Kimbrough. Exemplary documents: a foundation for information retrieval design. *Information Processing & Management*, 38:363–379, 2002.
- [3] H. H. Clark. *Common ground*. In *Using Language*. Cambridge University Press, Cambridge, 1996.
- [4] H. H. Clark and S. E. Brennan. Grounding in communication. In L. B. Resnick, J. M. Levine, and S. D. Teasley, editors, *Perspectives on Socially Shared Cognition*, pages 127–149. American Psychological Association, Washington DC, 1991.
- [5] G. W. Furnas, T. K. Landauer, L. M. Gomez, and S. T. Dumais. The vocabulary problem in human-system communication. *Commun. ACM*, 30(11):964–971, 1987.
- [6] S. R. Fussell and R. M. Krauss. 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.
- [7] M. D. Gordon. 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.
- [8] G. Jian and L. Jeffres. Understanding employees' willingness to contribute to shared electronic databases: A three dimensional framework. *Communication Research*, 33(4):242–261, 2006.
- [9] L. M. Markus. Toward a theory of knowledge reuse: Types of knowledge reuse situations and factors in reuse success. *Journal of Management Information Systems*, 18(1):57 – 93, 2001.
- [10] P. Rafferty. The representation of knowledge in library classification schemes. *Knowledge Organization*, 28(4):180–191, 2001.
- [11] S. Whittaker and J. Hirschberg. The character, value, and management of personal paper archives. *ACM Transactions on Computer-Human Interaction (TOCHI)*, 8(2):150 – 170, 2001.