Using Hierarchical Folders and Tags for File Management

A Thesis

Submitted to the Faculty

of

Drexel University

by

Shanshan Ma

in partial fulfillment of the

requirement for the degree

of

Doctor of Philosophy

March 2010

Dedications

This dissertation is dedicated to my mother.

Acknowledgments

I would like to express my sincerest gratitude to my advisor Dr. Susan Wiedenbeck. She encouraged me when I had struggles. She inspired me when I had doubts. The dissertation is nowhere to be found if it had not been for our weekly meetings and numerous discussions. I'm in great debts to all the time and effort that she spent with me in this journey.

Thank you to my dissertation committee members, Dr. Michael Atwood, Dr. Xia Lin, Dr. Denise Agosto, and Dr. Deborah Barreau, who have guided me and supported me in the research. The insights and critiques from the committee are invaluable in the writing of this dissertation.

I am grateful to my family who love me unconditionally. Thank you my mother for teaching me to be a strong person. Thank you my father and my brother for always being there for me.

I would like to thank the iSchool at Drexel University for your generosity in supporting my study and research, for your faculty and staff members who I always had fun to work with, and for the alumni garden that is beautiful all year round.

Thank you my friends in Philadelphia and my peer Ph.D. students in the iSchool at Drexel University. Those weekdays and weekends that we spent together contributed visibly and invisibly to the completion of my dissertation.

A special thanks to Marylin Huff's help in recruiting participants for my research. I would also like to acknowledge the help from all the Drexel students who participated in my research.

Table of Contents

List of Tables

List of Figures

Abstract

Using Hierarchical Folders and Tags for File Management
Shanshan Ma
Advisor: Susan Wiedenbeck, Ph.D.

Hierarchical folders have been widely used for managing digital files. A well constructed hierarchical structure can keep files organized. A parent folder can have several subfolders and one subfolder can only reside in one parent folder. Files are stored in folders or subfolders. Files can be found by traversing a given path, going through different levels of folders and subfolders. Folders can be moved, renamed, copied and deleted to serve the needs of the changing working environment. However, previous research has revealed several problems with hierarchical folder structures. One important problem is that users frequently have to turn to desktop search to re-find files. Tagging is the activity of applying users' own descriptors to digital objects, such as web pages, photos, and documents. Compared with traditional indexing which enforces a controlled vocabulary, tagging systems give users freedom in describing digital resources. We believe that tagging may have the potential to improve information navigation and information organization. This research aimed at exploring the possibility of incorporating tagging into the hierarchical folder structure for file management, especially for the process of file organization and file re-finding.

We studied users' behavior and preference of using three file management structures, a hierarchical folder structure, a tagging structure, and a hybrid structure with both hierarchical folder and tagging functionalities. We found that using tag alone or using folder alone generated similar results in file organization time, in file re-finding time and in answer correctness. Combining folders and tags resulted in longer file organization time but no improvement in file re-finding efficiency. The tagging structure required the least number of mouse clicks in the re-finding process among the three structures. The primary contribution of the study is a comparison of three file management structures for better organizing and re-finding files in the desktop environment. Advantages and disadvantages of each structure were revealed from the study. Users' preference among the three structures

was compared. Both quantitative and qualitative research methods were used in the research. This work will provide design implications for future file management tools.

Chapter 1: Introduction

File management refers to the process of accessing, managing, and maintaining files. In the digital world, a file refers to a collection of data or programs stored under a single identifying name (Oxford English, 2007). In digital environments, the most widely used file management system is the hierarchical folder system integrated within the computer operating system. The hierarchical folder system allows users to organize files into tree structures. Although computer itself and the operating system have gone through fundamental changes in the past few decades, the hierarchical folder structure stays almost unchanged (Marsden & Cairns, 2003).

A well constructed hierarchical structure can keep things organized. A parent folder often has several subfolders. A folder or subfolder can contain several files. Files are stored into folders and subfolders. Each file can be found by following a particular path that goes through different levels of folders and subfolders. Folders can be moved, renamed, copied and deleted to serve the needs of the changing working environment.

However, hierarchical folder structure is not a perfect solution for file management. Previous research has revealed problems with it. Users frequently have to turn to desktop search even if they thought they have saved their files in a memorable folder. New methods and designs have been introduced to help people maintain better control over digital documents. One category of research efforts come from the area of information retrieval. New desktop search systems have been developed to help users re-find files through new searching techniques. However, searching is not a silver bullet for re-finding. To re-find a desired file in one's personal computer, specific contextual information that is useful for re-finding files may not easily be captured in search keywords. Another category of research efforts attempt to develop alternative organization structures to hierarchical folders. The research on tagging is among this category of research efforts.

Social tagging is the activity of applying users' own descriptors to digital objects such as web pages, photos, blogs, and documents (Kipp & Campbell, 2006). Compared with traditional indexing, which enforcing a controlled vocabulary, social tagging systems give users freedom in describing digital resources. Social tagging systems such as Del.icio.us (www.del.icio.us), flickr (www.flickr.com), and CiteULike (www.citeulike.org) have been designed to enable sharing and managing web contents using tags. Practices from social tagging web

sites suggest that tagging has strong potential in changing information organization and management (Edmonds et al., 2004; Millen & Feinberg, 2006; Tennis, 2006). File management is an area that tagging may improve.

Tagging is claimed to be able to afford efficient information access and overcome the deficiencies of hierarchical file system (Bloehdorn et al., 2006). Tagging requires less cognitive load than categorizing, as suggested by Sinha's cognitive analysis of tagging (2005). It's easy to complete and provide immediate self and social feedback (Sinha, 2005; Tonkin, 2006). Meanwhile, tagging system has the flexibility that hierarchical system does not have. For instance, multiple tags can be applied to one information item and one tag can be applied to multiple information items.

However, problems could also occur to one's personal tagging system. Tagging system gives users freedom to index information with self-created terms. It also creates a large vocabulary that might cause problems when people try to re-find files with tags. Large number of self-created tags can grow to an extent that individuals forget and get lost in one's own self-created tags. Also, it happens frequently that people use different words to describe the same information item. Tagging inconsistency makes tags hard to manage.

We expect that a hybrid system with both hierarchical folders and tags might solve the deficiencies of each structure when used alone. Files are organized with both folders and tags. File systems with tagging functions have already been proposed and developed previously. However, most of such efforts are still in preliminary stage (Bloehdorn, et al., 2003; Taggtool, 2006). There is limited empirical research investigating this problem. Whether or not tagging can complement the hierarchical folder system for file management is still a myth. We are not only interested in comparing hierarchical folders and tags for file management, but also in whether tagging can complement folders for file management. The high level research goals of this dissertation are:

- To gain better understanding of the advantages and disadvantages of using hierarchical folders and tags for file management.
- To explore the possibility of incorporating tagging into the hierarchical folder structure for file management.
- To provide design insights for future file management tools.

Chapter 2: Literature Review

2.1 The Research of Personal Information Management (PIM)

In order to gain better understanding of the nature of file management, we looked deeply into the literature of personal information management. Personal information is the kind of information people keep for their own use (Lansdale, 1988), such as books we own, files we have in our cabinet, documents we save on our hard disks, and emails. Personal Information Management (PIM) refers to "the methods and procedures by which we handle, categorize, and retrieve information on a day-to-day basis" (Lansdale, 1988). Jones and Bruce (2005) defined PIM activities into three groups: keeping, finding/re-finding, and M-level activities (mapping or maintenance and organization). Each group of activities affects a different aspect of personal information space. Keeping activities affects information input; Finding and re-finding activities affects information output. M-level activities affect information storage which includes information maintenance and organization.

Barreau (1995) described five functions of a PIM system: acquisition, organization, maintenance, retrieval, and output. In her description, acquisition of information refers to the process of adding information to people's personal information collection. Organization refers to the process of arranging information in personal information space. Maintenance refers to the process of updating what has been saved to cope with the changing working environment. Retrieval refers to the process of finding information for reuse.

PIM activities take places in both physical environment and digital environment. Previous PIM researches in physical environment were conducted to help design new PIM tools in the digital environment (Bondarenko et al., 2005; Kaye et al., 2006; Lansdale, 1988; Malone, 1983; Whittaker et al., 2001;). In particular, the importance of physical paper in the modern digital environment for facilitating personal information management was discussed intensively. Malone's study on how people organize physical desks (1983) revealed many interesting findings that impacted PIM research in the digital environment. He pointed out that it is important for desk organization to support reminding functions. Malone also talked about how cognitive difficulty of categorizing information can influence the way people organize their physical desks. Malone documented two kinds of people with paper organizing: filers who do filing, which means putting files into

structured folders, and pilers who do piling, which means information is left unprocessed and distributed around in the working space.

In the digital environment, a large body of research looked at personal information management across various technical applications (e.g. Alvarado et al., 2003; Bergman et al., 2006; Boardman & Sasse, 2004; Jones, 2004). They tried to look into how people carry out personal information management activities in the digital space as a whole. Since users are usually working on a wide range of tools and technologies, information fragmentation is a big problem that users are facing. Information fragmentation refers to the situation when users store and retrieve relevant information across different applications. Such situation causes problem when users try to re-find information that resides in different collections in different formats. Many researchers discussed this issue in their research work (Bergman et al., 2006; Bellotti & Smith, 2000; Jones, 2004). Boardman (2004) made effort towards integration of multiple applications.

Besides research on information fragmentation, extensive study has been done on individual technical tools and applications. Email is one of the most widely discussed digital tools for PIM. Previous research (Bellotti et al., 2002; Berghel, 1997; Ducheneaut et al., 2001; Whittaker et al., 1996) studied how people struggled with the ever-growing size of email inbox. In particular, Whittaker et al. (1996) used the word “email overload” to describe the situation that email has been used for purposes beyond asynchronous communication, such as task management and personal archiving and filing. Similar to Malone (1983)’s differentiation between filers and pilers, Whittaker identified three categories of people based on the strategies they used for handling email overload: No filers; frequent filers; and spring cleaners. No filers made no use of folders. Frequent filers made frequent use of folders and tried to minimize the numbers of inbox messages. Spring cleaners made extensive use of folders and they had large overloaded inboxes. Normally they tried to clean-up their email boxes every 1to 3 months.

Study in organizing bookmarks and preserving previously found web pages is a another big category in the field of PIM research. Research under this topic discussed about the differentiation between finding information and re-finding information from the web (Capra et al., 2004; Capra & Perez-Quinones, 2006; Teevan, 2004); finding and keeping behavior with web information (Jones et al., 2000; Jones et al., 2001); and the management of bookmarks for web information preservation (Abrams et al., 1998).

2.2 File Management with Hierarchical Folders

Hierarchical folder structure has been used widely for managing information in the digital environment ever since the first computer was invented. Rigorous research has been done looking into folder's role in file management (Akin et al., 1987; Barreau and Nardi, 1995; Bondarenko et al., 2005; Henderson, 2004). Akin et al. (1987) studied the structure of directory space under UNIX operating system. It was among one of the earliest study on hierarchical structures. Barreau and Nardi (1995) studied users' filing management under different operating systems, including DOS, windows and OS/2, and Macintosh. Berreau and Nardi (1995)'s research triggered a series of debates regarding users' browsing and searching behavior in the process of finding files (Barreau and Nardi, 1995; Barreau and Nardi, 1997; Fertig et al., 1996). Our following discussion on file management derives from this aspect of PIM research. This section will focus on research and practices in using hierarchical folders for file management. We'll discuss the concepts of file, folder and file management. This section will extensively cover problems with hierarchical folder systems that previous research has discovered.

2.2.1 The Concept of File, Folder, and File Management

In the digital world, a file refers to a collection of data or programs stored under a single identifying name (Oxford English, 2007). A file can be presented in different formats: a bit of text, a graphic image, an audio record, or anything else. The size of a file can vary from a few bytes to multiple gigabytes. A folder or a directory refers to a container that contains a list of files and subfolders. Usually folder name describes what kind of files and subfolders that the parent folder contains. A folder can either be empty or contain thousands of files.

At the lower level, files are streams of bytes, and folders/directories contain file names that link to the content of the files. Different file system may have different mechanisms in storing files and directories to hard disks. However, at the higher level, we see files as individual objects and folders as containers where files reside, just as what we see physical files and physical folders.

File management refers to the process of accessing, managing, and maintaining files. In the digital environment, the most widely used file management system is the hierarchical folder structure integrated within computer

operating system. The hierarchical file system allows users to organize files into a tree structure. A well constructed hierarchical structure can help us keep things organized.

Barreau (1995)'s categorization of PIM activities: acquisition, organization, maintenance, retrieval, and output, makes good sense in our research on file management. However, we use "re-finding" instead of "retrieval" in this dissertation, in order to differentiate from retrieval as in a searching system. As many researchers have stated in their research, personal information is stored and organized to be found again for later reuse (Akin et al., 1987; Barreau & Nardi, 1995; Lansdale, 1988). The problem of file management is essentially the problem of efficient organization and effective re-finding. In our research, we mainly focus on the process of file organization and the process of file re-finding.

2.2.2 Organizing Files with Hierarchical Folders

People have been using hierarchical folder for file management for a long period of time, even before the digital era. Library collections are organized with a hierarchical classification system based on their subject. In the digital environment, operating system has gone through many generations of transformation, but hierarchical folder system stays barely changed.

A well constructed hierarchical structure can keep things organized. A parent folder can have several subfolders and one subfolder can only belong to one parent folder. Files are stored into folders. Each file can be found through a particular path that goes through different levels of folders and subfolders. Folders can be moved, renamed, copied and deleted to serve the needs of the changing working environment.

Usually, users would categorize files and create folders based on the content of the file. Henderson (2005) looked at the types of folders people create and how they organized them into hierarchies. She found that most folder names represented the genre, task, topic or time dimension of the documents that the folder contained. The context in which a file was acquired, created or used would largely affect how it was classified, stored, and later retrieved (Barreau, 1995). Kwasnik (1989) and Barreau (1995) found that contextual information such as situational attributes, document attributes, disposition, order/ scheme, time, value, and cognitive state were used as key factors in the organization of materials.

2.2.3 Re-finding Files with Hierarchical Folders

Akin et al. (1987) found that in UNIX system, users tended to first try to open directories based on their recall of the files that they wanted to find, and then recognize them in those directories. This finding was reconfirmed by later studies in other operation systems. Barreau (1995) found that users usually used the information they can recall about a file to roughly decide where they should go to get access to the file. Then they browsed through a list of files and found the desired file by recognizing it from the list of files. Barreau called this behavior "location-based search". Although Fertig et al. (1996) argued that "location-based search" is no different from a user controlled keyword search; later studies suggested that they are more different than they are similar. Alvarado et al. (2003) found that people prefer to "use contextual information as a guide in navigating locally in small steps" rather than directly using keyword search. Boardman & Sasse (2004) reported similar findings to Alvarado et al.'s finding. They found that participants preferred browsing over searching. To summarize these previous discussions, in the context of desktop file management, instead of using a direct keyword search, users prefer to first "recall" information about the file they want to find, roughly decide where they can possibly find the target file, and then browse through the list of files in the location to "recognize" the desired file.

2.2.4 The Role of Memory in File Organization and File Re-finding

Memory plays a significant role in the process of file organization and file re-finding. How users organize files can affect users' memory about the files. What users recall about the file decides how they will to find it. As Lansdale (1988) concluded, "the ability to recall information depends upon a critical relationship between how the information is held in memory and what we are thinking about when we are trying to retrieve it." Generally speaking, users can partially or fully recall the following features of a file: location, type or format, file name, title, size, time, keywords, visual elements, associated events, links, and action (Blanc-brude & Scapin, 2007).

In an ideal situation, users recall just enough features about the desire file that allow them to go exactly where they saved the file and re-find it. In reality, users struggles that what they remember about the desired file is not sufficient for them to re-find the file successfully. They have to rely on the organization system's reminding function to guide them to the file that they are looking for. Barreau & Nardi (1995) observed that "location-

based finding" provided reminding function when users navigate through different files. Jones et al. (2005) found out that folder itself can serve as a kind of information for reminding during the re-finding process.

2.3 Hierarchical Folders: What Goes Wrong?

Previous research has revealed problems of using hierarchical folder for file management, mainly from four aspects: categorization ambiguity and cognitive difficulty, file invisibility caused by the deep file hierarchy, lack of contextual reminding functions, and lack of regrouping and restructuring capability. This section will discuss these problems.

2.3.1 Categorization Ambiguity and Cognitive Difficulty

Dumais & Landauer (1983) discussed two reasons why information can not fall into neat categorizations. First, category name by nature can never be used unambiguously. Second, information in real world situations always falls into "overlapping and fuzzy categories". Such ambiguity caused cognitive difficulty in categorizing as reported by the following researches. Malone's (1988) research pointed out that the cognitive difficulty user experienced in classifying information was one of the main challenges users faced in PIM. Whittaker et al. (2001)'s survey and interview again reported cognitive difficulties experienced by users in deciding how a given document should be categorized. To relieve the problem of cognitive difficulty, Malone (1988) suggested multiple classification and deferred classification. Multiple classification structure allows the same document to be put into several categories. Deferred classification allows users to defer classification as physical piles which don't require explicit classification. Quan & Bakshi (2003) did a controlled study to compare a multiple categorization system and a regular hierarchical folder system. Multiple categorization system received positive results in the experiment.

2.3.2 Deep Hierarchy Makes Files Invisible

In a traditional hierarchical folder system, a file can only reside in one and only parent folder, and not in higher-level folders (Bloehdorn, et al., 2006). When opening a folder, files that are contained in its subfolders remained unseen. Boardman & Sasse (2004) reported an average depth of 3.3 in hierarchical folder systems. It means that if users decide to use "location-based search" as the strategy to re-find a file, they have to follow through the

exact 3.3 level saving path to get access to the file. Balter (2000) suggested that extensive and deep filing structure is not as efficient as a flat and simple file structure. However, a flat and simple file structure will result in large number of files within folders, which in turn would make the recognition process harder.

2.3.3 Lack of Contextual Reminding Functions

The contextual information that users recall about the file would decide how it will be found. Bondarenko & Janssen (2005) discussed that it is hard to attach contextual information to files when putting files into a digital file folder. Visual cues and textual cues which are available in physical PIM environment are no long applicable in the digital hierarchical folder environment. Besides the file format which is reserved by the file extension, the only contextual information saved associated with the file is the levels of folders that the file belong to. However, if the file is moved to another location and no longer reside the previous parent folder, such contextual information will no longer be available.

2.3.4 Lack of Regrouping and Restructuring Capabilities

The rigid hierarchical structures are hard to maintain and keep up with the evolving desktop information space. Bondarenko & Janssen (2005) witnessed the shortcomings of current folder system in that it is hard to regroup and restructure files to meet the needs of the changing task flow. A more flexible document management system that can easily regroup and restructure files based on task environment is needed.

2.4 Solutions for Better File Management

2.4.1 Desktop Search

Better search systems try to help users re-find files through advanced technical approach. However, search is not a silver bullet. Users usually want an “exact match” when they find things from their own collection. In other words, a high “precision” is desired when trying to find files from one’s personal computer. Most of the time users know exactly what they are looking for in file management even if they don’t remember exactly what the file is. In order to get the search result returned exactly what they expected, users need to have more restrictions in search terms to get the results. Usually this means more work from users’ side. As we discussed

previously, users tend to choose a "location based search" or a "recall and recognition" approach instead of a "keyword based search" in the beginning of the re-finding process.

Similar to web search engines, most current desktop search systems only support keywords search (Cutrell, 2006). New desktop search systems have been developed to support re-finding from personal file collection. "Stuff I have seen" (Dumais et al., 2003) is a prototype designed by Microsoft for personal file retrieval. The tool indexed all the information that users have seen in their digital working space. Users were able to use contextual cues for search, such as when the item was encountered, who the author is, how big it is, and etc. Phlat was another system designed to support personal information search. Phlat "merges search and browsing through a variety of associative and contextual cues" (Cutrell et al., 2006).

2.4.2 Beyond Search

Other research efforts in PIM attempt to look for replacements to hierarchical systems. Using the spatial metaphor for file management is one area of these research efforts. There are debates over whether using the spatial metaphor for digital file management is effective and no decisive conclusion has been reached. Based on experiment results, Jones & Dumais (1986) inferred that spatial representation of file organization may not be as effective as people thought. Cockburn & McKenzie (2002) also found that users' performance of organizing and finding documents deteriorated as higher dimension added from 2D to 3D, both in physical and virtual systems. However, there are also studies proving that a spatial metaphor might be more helpful compared with the symbolic file system. Mander et al. (1992) designed a system to support casual information organization using the "pile" metaphor. Robertson et al. (1998) created a system called "Data Mountain" that "allows users to place documents at arbitrary positions on an inclined plane in a 3D desktop virtual environment". Bauer et al. (2005) designed a system called "Dynapad" that facilitated "visual access to and spatial organization of digital photo collections and personal libraries of PDF files".

Other than the spatial metaphor, a series of research has been done by Freeman et al. (1996a, 1996b, 1996c) on dynamical file organization with a system called "Lifestreams". Lifestreams used a time-ordered stream of documents to replace traditional folders and files. A lifestream is "a time-ordered stream of documents that functions as a diary of your electronic life". Presto, developed by Dourish et al. (1999), was a prototype

document management system that provided rich interaction with documents through various documents attributes. Voida & Greenberg (2009) saw the potential of using Wikipedia to keep things organized on the web. They proposed a hybrid system for annotating files that draw upon the strength of both the hierarchical folder system and a wiki system.

2.5 Tagging

2.5.1 Social Tagging

Social tagging, also called folksonomy or collaborative tagging, refers to user generated taxonomy that is used to categorize and retrieve web content such as web pages, photographs, and web links (Kipp & Campbell, 2006). Tags are usually one word descriptors. People use tags to label web content for information organization and information sharing purpose on web sites such as Del.icio.us (del.icio.us), Flickr (www.flickr.com), and CiteULike (www.citeulike.org). Del.icio.us is a social bookmarking web site that allows individuals store, share and discover web bookmarks. Flickr is a web site that allows individuals to share photos. CiteULike is a web site that helps researchers to share, store and organize academic papers.

How does social tagging work? Take Del.icio.us as an example, upon visiting, users are asked to install a bookmark extension based on what kind of browser they are using. Two buttons will then show up in the browser after the installation: one marked with "Del.icio.us", the other one marked with "tag". When users see a web page that they would like to add to their bookmark collection, they click the "tag" button to save the page. They can assign tags to the page by filling out a form. The saved web page will always be available under users' personal account on Del.icio.us, which can be accessed by clicking on the Del.icio.us button in the browser. Tags created are not o when saving the web page are not only used for one to organization purpose, but also used for bookmark sharing across the entire Del.icio.us system. For example, when clicking on the tag of "travel", one will be presented with all the web pages that are tagged with "travel" by users of Del.icio.us.

Compared with a controlled vocabulary, which usually forms a hierarchical structure, tags create a free-structured flat namespace. There have been ongoing research and discussion on using tagging for knowledge organization and resource discovery. Feinberg and Millen (2005)'s paper about using tagging for improving social navigation introduced a design of a social bookmarking service software called "dogear". The software

enabled users to assign tags to web pages and organize web pages using tags. Tennis (2006)'s paper presented similarities and differences between social tagging and subject cataloguing. However, only limited number of peer reviewed scholarly research are available, with a lot others reside in personal blogs and web discussions. Web discussions, especially web blogging communities (Macgregor & McCulloch, 2006), seem to attract more enthusiastic debates and arguments over the potential of using tagging for information organization and navigation. For instance, Shirky (2005) has a series of blog articles suggesting that collaborative tagging will supersede controlled vocabulary because of its inclusive, free, and participatory nature.

2.5.2 Using Tagging for File Management

We believe that there are potential advantages of using tags for file organization. First, tagging is believed to be cognitively easier than categorizing (Sinha, 2005). Classifying and categorizing information requires cognitive effort. Real life information is ambiguous to be classified into absolutely neat hierarchical structures. Tags are believed to require less cognitive load than categorization because tags could be any text that is associated with the information. They do not have to be a high level generalization or categorization. Second, hierarchical systems require users to follow the exact saving path to get access to the desired file. With tagging, the target file can be accessed with any associated tag. Third, users can give contextual tags to files, such as "important", which can serve as reminders later in the process of re-finding. Fourth, multiple tags can be assigned to a file. Since there is no parental and sibling relationship between tags, files can be easily regrouped from different perspectives. Research has been done attempting to explore the possibility of using tags for file management. For example, a software tool called "Tagg" was designed to help users manage personal files with tags (www.taggtool.com). Tagg allows users to attach tags and descriptions to files.

However, there are potential problems with social tagging; problems could also occur to one's personal tagging system. First, large number of self-created tags can grow to an extent that individuals forget and get lost in one's self-created tags. Second, people may use different words to describe the same information, which would make it hard to find files only with tags. Third, the way that tags are presented is a problem. Currently tag cloud is used widely in social tagging systems. Basically, tags cloud is a set of alphabetically presented tags, with the

most frequently used tags emphasized in a larger font size. Whether tag cloud is a useful mechanism for a personal tagging system is not clear.

As discussed previously, both hierarchical folder systems and tagging systems have their own strengths and drawbacks. We expect that a hybrid system with both hierarchical folders and tags will partially solve the deficiencies of each structure when used alone. There has been research work aiming at integrating tagging into the hierarchical file systems. Bloehdorn et al. (2006) proposed TagFS. In TagFS, files are annotated with RDF metadata so that "file system operations are mapped to manipulations of tags on information objects". Civan et al. (2008) saw the advantages and disadvantages of using either hierarchical folders or tags for managing information. They conducted an exploratory study into how people use folders and tags differently for managing emails.

In our research, we ask three research questions as listed below.

RQ1: To what extent does tagging support file management, especially for file organization and file re-finding?

RQ2: To what extent does combining tagging and folder system support file management, especially for file organization and file re-finding?

RQ2: What are users' preference among the three structures, the hierarchical folder structure, the tagging structure, and a hybrid structure?

Chapter 3: Methodology

This chapter describes the methodology used in the dissertation – a controlled laboratory study comparing the process of file management using three different structures: a folder structure, a tagging structure, and a hybrid structure with both folder and tag functionalities. The study had two experimental sessions, scheduled one week apart. In the first experimental session, participants were asked to save and organize 60 files using an assigned organization structure. In the second experimental session, participants were asked to re-find files from the file collection that they saved in the first session.

3.1 Participants

Participants who were interested in information and technology were recruited from Drexel University. Fifty-one participants were enrolled in the experiment, but two participants could not return for the second session because of schedule conflicts. Forty-nine participants, 35 men and14 women, completed the study. The distribution of the participants is shown in table 1.

Table 1. Participants Demographic Information

	Male	**Female**	**Total**
The Folder Group	11	5	16
The Tagging Group	12	4	16
The Hybrid Group	12	5	17
Total	**35**	**14**	**49**

Forty-six participants were between the age of 18 and 25. Three participants were between 25 to 35 years old. In terms of academic majors, 39 participants were from the Department of Engineer, the remaining participants from the College of Information Science and Technology, Business/Finance, Interior Design and Nursing. Forty-seven participants were undergraduate students and two participants were master students. Forty-six participants reported that they had more than five years experience with computer and three had 1to 5 years experience with computer.

3.2 Materials

3.2.1 Zotero

Zotero is a software tool for gathering and organizing sources including web pages, citations and images. It was designed as an extension to the Firefox web browser. We chose this tool for our experiment because it supports both hierarchical folders and tags under one interface. Zotero was customized into three versions for our experimental purpose: a version for the folder group that supports hierarchical folders for file management; a version for the tagging group that supports tags for file management; and a version for the hybrid group that supports both folders and tags. The advantage of using Zotero is that participants in all three groups have the very same basic interface.

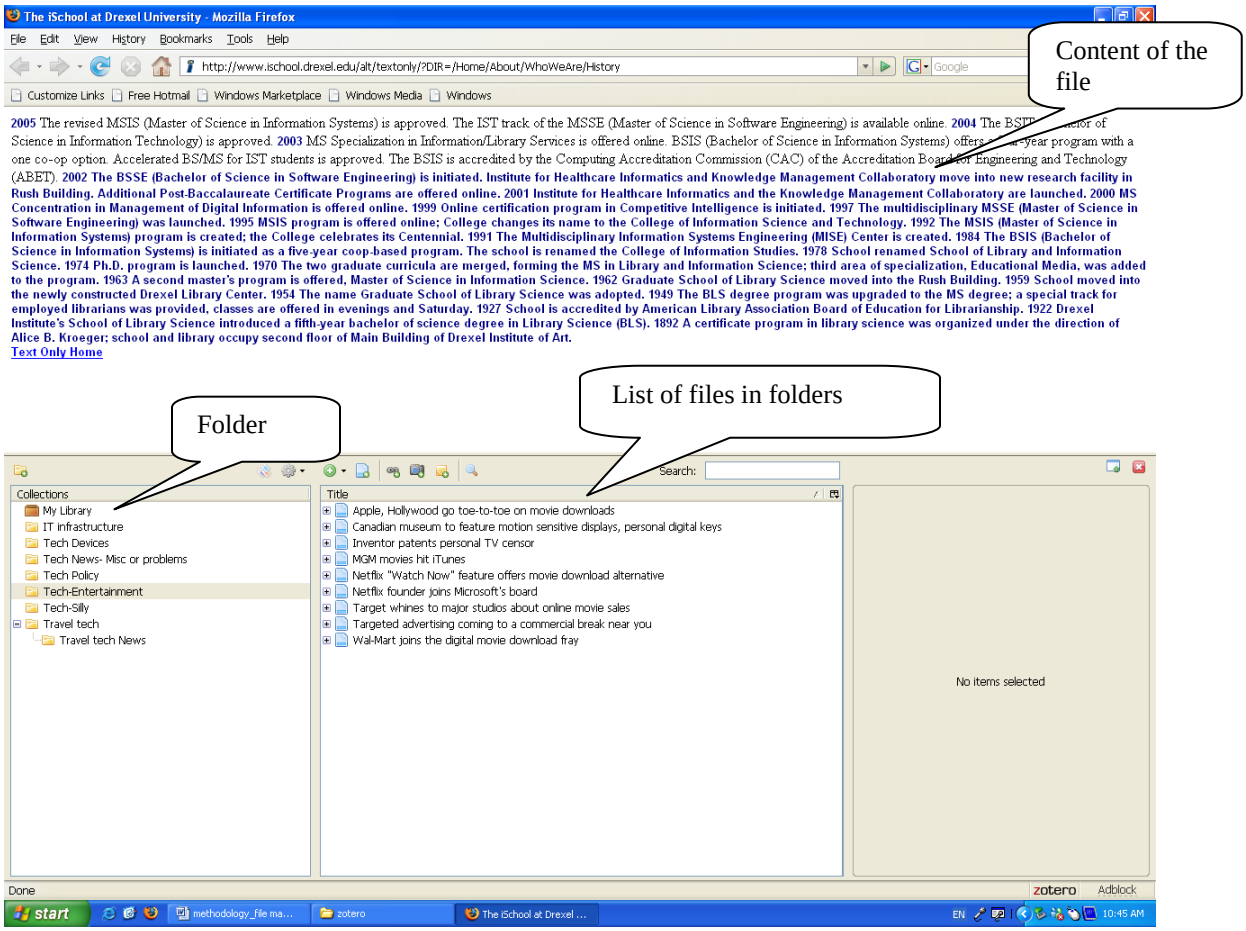


Figure 1. The Zotero Interface for the Folder Group

The Zotero interface for the folder group is shown in figure 1. The window is divided into two sections: the upper level of the window shows the content of the current file in the format of a web page. The lower level of the window is Zotero's working space. It allows participants in the folder group to organize files into hierarchical folders. Zotero's working space is divided into three panes: left, middle and right. The left pane shows the organization structure developed by the participants. The middle pane shows a list of files that participants have saved. The right pane is not used for the folder group. When participants finish reading the file content in the upper level of the window, they create a folder with appropriate folder name on the left pane, and save the file to the folder. The file is then listed in the middle pane.

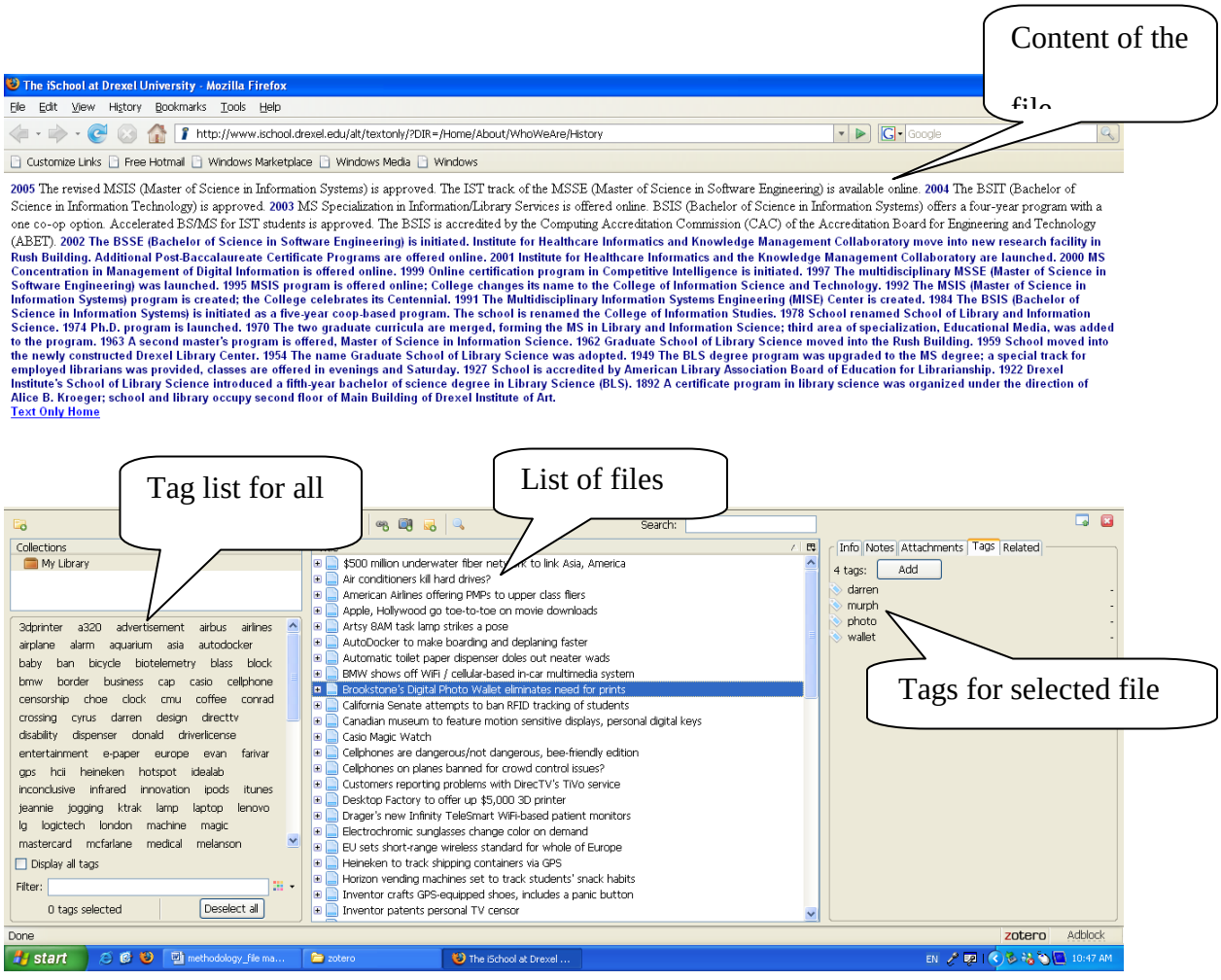


Figure 2. The Zotero Interface for the Tagging Group

Figure 2 shows the Zotero interface for the tagging group. The window is divided into two sections as in the folder group: the upper level of the window shows the content of the current file in the format of a web page. The lower level of the window is Zotero's working space. It allows participants in the tagging group to organize files using tags. Zotero's working space is divided into three panes: left, middle and right. The left pane shows the complete list of tags created by participants. The middle pane shows the files that participants have saved. The right pane shows tags belong to any selected file. When participants finish reading the file content in the upper level of the window, they save the file and the file is then listed in the middle pane. Participants then create relevant tags on the right pane and the newly created tags will be added to the complete tag list on the left pane.

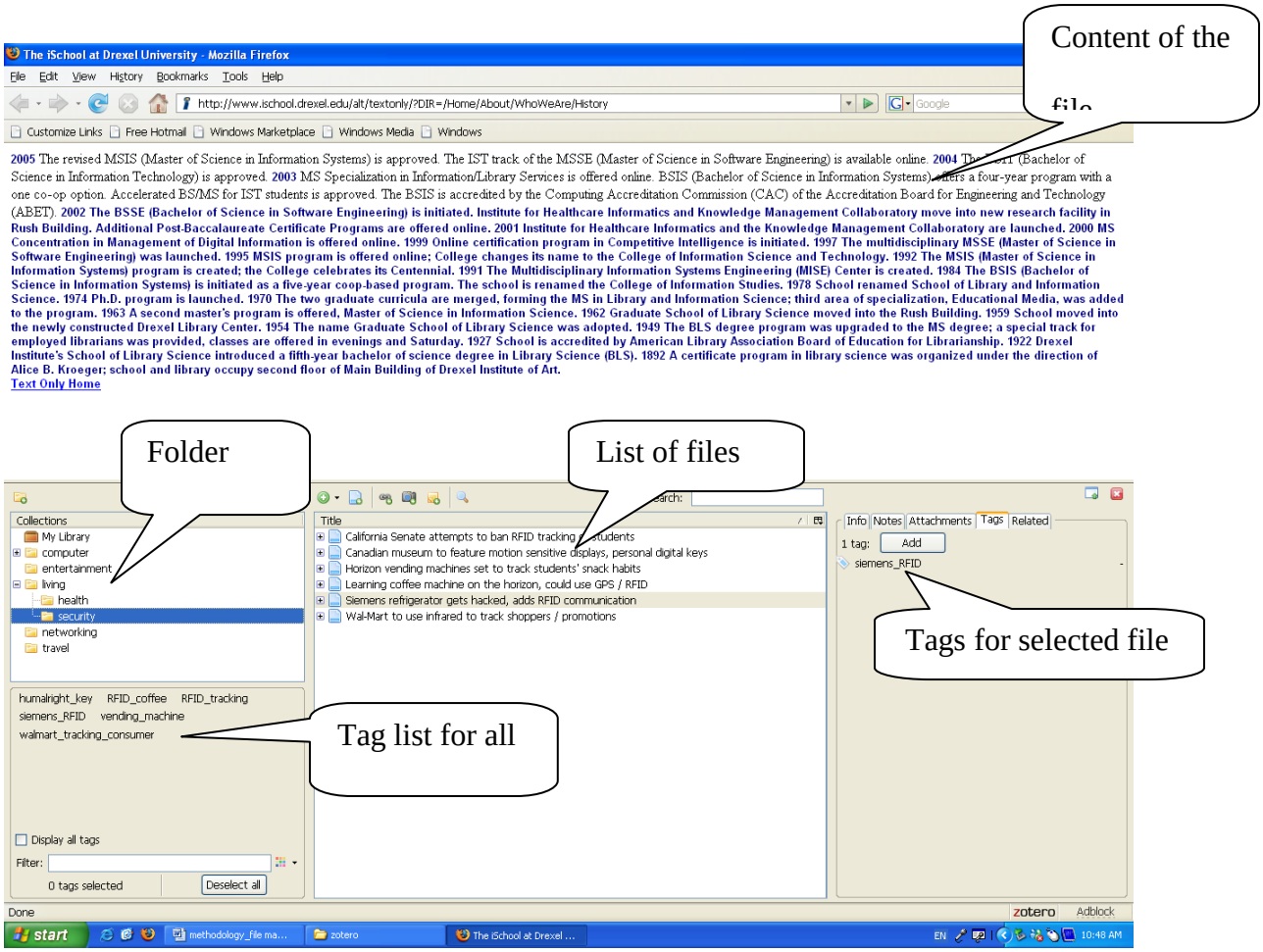


Figure 3. The Zotero Interface for the Tagging Group

Figure 3 shows the Zotero interface for the hybrid group. Similar to the other two groups, the window is divided into two sections: the upper level of the window shows the content of the current file in the format of a web page. The lower level of the window is Zotero's working space. It allows participants in the hybrid group to organize files using both hierarchical folders and tags. Zotero's working space is divided into three panes: left, middle and right. The left pane shows the organization structure including folders and tags created by participants. The middle pane shows the files that participants have saved. The right pane shows tags belong to any selected files. When participants finish reading the file content in the upper level of the window, they create a folder with appropriate folder name on the left pane, and save the file to the folder. Participants then create relevant tags on the right pane and the newly created tags will be added to the complete tag list on the left pane.

3.2.2 Tutorial

Although participants were familiar with folders, they were new to the Zotero interface. Therefore, before the experiment started, participants were given a written tutorial to learn about how to use Zotero for file management. The tutorial contained step-by-step instructions of how to use Zotero. The tutorial explained how to create a folder and how to assign a tag in the Zotero interface. The tutorial also included hands-on practice so that participants can interact with Zotero as they went through the tutorial. Three versions of the tutorial (Appendix A-1 to A-3) were developed so that participants in each group only learned about the functions that were relevant to their tasks. Also, participants who worked with tags (the tagging group and the hybrid group) were given a tutorial on tags (Appendix A-4). This tutorial included a short introduction about what a tag is and a list of tags from Flickr (www.flickr.com) as examples.

3.2.3 Article Collection

Participants worked on an article collection composed of 60 files, which were chosen from a blog site called Engadget (www.engadget.com). Engadget is a web magazine with daily coverage of news on gadgets and consumer electronics, with topics such as cellphones, laptops, household devices, wearable devices, wireless technology, network technology, high tech equipped transportation, and new media technology. The chosen articles were all one paragraph long and contained no more than 400 words. Please refer to appendix F for a complete list of the 60 articles.

We chose the number of 60 for two reasons. First, we needed a large number of articles to make the experiment realistic, in terms of organizing and re-finding the articles. A smaller number of articles would have been easily manageable regardless of the organization scheme. Second, more than 60 articles would take so much time that participants are likely to become bored and frustrated. Our pilot study suggested that 60 articles were appropriate in that all participants completed the organization session without showing symptoms of boredom or frustration.

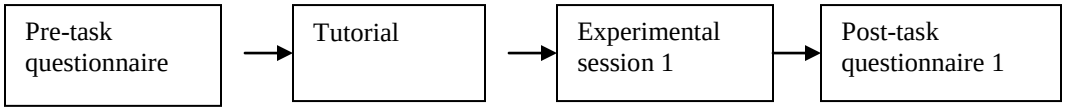
3.2.4 MORAE

Morae (www.morae.com) is a software tool that was used in the experiment to record the interaction between participants and the experiment interface during the two experimental sessions. Morae has three components: Recorder, Remote Viewer and Manager. Morae Recorder was used to capture and synchronize video and data input, creating a digital record of system activity and user interaction. Morae Remote Viewer allowed the researcher to monitor the Recorder over a network connection. Morae Manager was used for data management.

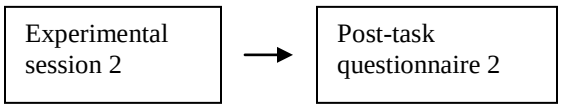
3.3 Experimental Procedure

The study consisted of two parts. The first part included a pre-task questionnaire, a tutorial, experimental session 1, and post-task questionnaire 1. The second part included experiment session 2 and post-task questionnaire 2. The two parts were scheduled one week apart. Each participant completed the two sessions individually.

Part 1: File Organization



Part 2: File Re-finding



3.3.1 Pre-task Questionnaire

Demographic information of the participants was collected from the pre-test questionnaire (Appendix C-1). The questionnaire included questions asking about gender, age, academic major, educational degree and computer experience. The pre-task questionnaire also asked the participants to rate their previous experience with file management and tagging, using a five item rating scale, ranging from strongly disagree, disagree, neutral, agree to strongly agree.

3.3.2 Tutorial

After participants filled out the pre-task questionnaire, they were given the tutorial corresponding to the group they were assigned. Participants read the tutorial, followed the steps, practiced on the computer, and learned about the functions that they would use later in the experimental sessions. The researcher sat besides the participant and was ready to help if needed. Participants were encouraged to ask any questions during the tutorial. The tutorial took approximately 20 minutes to finish.

3.3.3 Experimental Session 1: File Organization

After participants completed the tutorial, they began experimental session 1, file organization. Participants were given a task description at the beginning of the session (Appendix B-2). They were asked to read through the article collection one by one and create an organization scheme from scratch, using hierarchical folders, tags, or both, based on which group they were assigned. Participants in the folder group were allowed to create as many folders as they wanted; however, a file was not allowed to be put into more than one folder. The reason for this restriction was that we observed in the pilot study that some participants intentionally put a file into most or all folders so that they could later retrieve the file without effort. Participants in the tagging group were allowed to assign as many tags as they wanted. Participants in the hybrid group were allowed to create as many folders and assign as many tags as they wanted. Participants were allowed to change the organization scheme at any time as they read through the article collection. Participants' interaction with Zotero was recorded using Morae recorder, for example, moving an article to different folders; creating folders for the article; merging folders; deleting folders; adding and deleting tags. The researcher monitored the participant's activity across the entire session.

Participants were timed, but they were advised to spend as much time as they needed to organize the files. Participants' time for completing the experimental session ranged from approximately one hour to nearly two hours.

3.3.4 Post-task Questionnaire 1

When participants completed organizing the 60 articles, they were asked to fill out a questionnaire, the post-task questionnaire 1(Appendix C-2). It included six to eight statements (varying among groups) that participants rated using a five item rating scale, ranging from strongly disagree, disagree, neutral, agree to strongly agree. The statements described participants' perception of the performance of the organization session. In addition, participants answered three open ended questions: please describe the strategy you used to organize the articles; what are the functions you would like the system to have while it doesn't; what are the problems you experienced while organizing the articles, if any.

Participants spent approximately 10 minutes to complete the questionnaire.

3.3.5 Experimental Session 2: File Re-finding

One week later, participants returned and began the second experimental session, file re-finding. They were given task descriptions at the beginning of the session (Appendix B-3). Participants were asked to re-find several articles from the article collection that they had saved previously.

They were given four different re-finding tasks sequentially, each providing different clues based on which participants had to re-find the target files. In the first task, five article titles were presented. After they finished, five article summaries were presented in the second task. Each summary was one sentence long. Special attention was paid to the wording of the summaries so that the summaries did not contain any keywords from the title. In the third task, five questions were presented. No keywords from the article titles were used in the wording of the questions. One article was relevant to each question. The task was considered finished once participants thought they had found the relevant article. In the fourth task, three questions were presented. In this case, multiple articles were relevant to each question. No keywords from the article titles were used in the

wording of the questions. The task was considered finished once participants thought all the relevant articles had been found. Table 2 shows an example to each of the re-finding tasks.

Table 2. Example of the Re-finding Tasks

Task	Types of tasks	Example
1	Article titles	LG Philips announces A4 color e-paper.
2	Article summaries	This device does more than just telling the time.
3	Questions asking for single articles	What's the new way that motorists in Ohio pay the toll?
4	Questions asking for multiple articles	Technology enhances the way people do exercises and sports. Please find all the relevant articles.

For each re-finding task, participants read the clues, navigated through the organization scheme they created in the previous session, and tried to find the correct target files. When participants accessed the correct article, they typed the article number in an appropriate box in the Answer Sheet (Appendix D). The Answer Sheet was a Word document that was always open on the screen. In cases when participants failed to find an article, they left the corresponding box open and moved on to the next task.

Participants completed the re-finding tasks in sequence. Participants' interaction with Zotero was recorded with MORAE Recorder. The researcher monitored the participant's activity across the entire session. A help card (Appendix E) was available in case the participants forgot certain functions of Zotero. They were asked to finish the tasks as quickly as possible and the re-finding process was timed. Overall, participants spent approximately 30 minutes in completing the file re-finding time.

3.3.6 Post-task Questionnaire 2

When the second experimental session ended, participants filled out another questionnaire, the post-task questionnaire 2 (Appendix C-3). It included five to seven (varying among groups) statements that participants rated using a five item rating scale, ranging from strongly disagree, disagree, neutral, agree to strongly agree.

The statements described participants' perception of the performance of the file re-finding session. In addition, they answered six open ended questions: please describe how you found the articles in task 1, task 2, task 3, and task 4; what are the functions you would like the system to have while it doesn't; what are the problems you experienced while completing the task, if any.

It took about 10 minutes to complete the questionnaire. After this session, participants received the compensation for their participation.

Chapter 4: Quantitative Data Analysis, Results and Discussion

This chapter reports and discusses the results from quantitative data analysis. As stated in the previous chapter, the research was motivated by three research questions.

RQ1: To what extent does tagging support file management for file organization and file re-finding?

RQ2: To what extent does combining tagging and folders support file management for file organization and file re-finding?

RQ 3: What are users' preference among the three structures, hierarchical folders, tagging, and the hybrid structure?

Comparisons among the three groups were carried out to determine how different organization structures affect the way people organize and re-find digital files. In this chapter, comparisons are mainly conducted on quantitative variables: file organization time, number of tags/folders created, file re-finding time, number of mouse clicks, answer correctness, and user perception. One-way ANOVA was used for all the comparisons in this chapter.

4.1 How Were the Three Groups Different in the File Organization Session?

4.1.1 File Organization Time

The task completion time for the file organization session was measured as starting from the time participants initiated the session by clicking on the first article title in the index page to the time they finished saving the last article and revised the organization scheme as desired. File organization time was captured by the screen recordings. For most participants, the file organization session was over when the participants finished saving the 60 articles. However, seven participants spent extra time in revising the organization schemes after they saved the last article. Consequently, we used two measurements of the file organization time: the total organization time with the extra revising time and the total organization time without the extra revising time.

4.1.1.1 File Organization Time with Extra Revising Time

A comparison of file organization time with extra revising time was conducted across the three groups. Table 3 shows the mean and standard deviation of file organization time (in seconds) with extra revising time for the three groups. ANOVA showed that there was a significant difference, $F(2, 46) = 7.61$, $p<.01$. Post-hoc Tukey's HSD test showed that file organization time with extra revising time was significantly longer in the hybrid group than the folder group, $p<.05$. No significance was found in other pair comparisons. Figure 4 shows the mean value of the total file organization time for the three groups.

Table 3. File Organization Time with Extra Revising Time (in seconds)

	Mean	**Standard Deviation**
The Folder Group	3352.69	1039.77
The Tagging Group	4021.62	756.00
The Hybrid Group	4717.94	1164.68

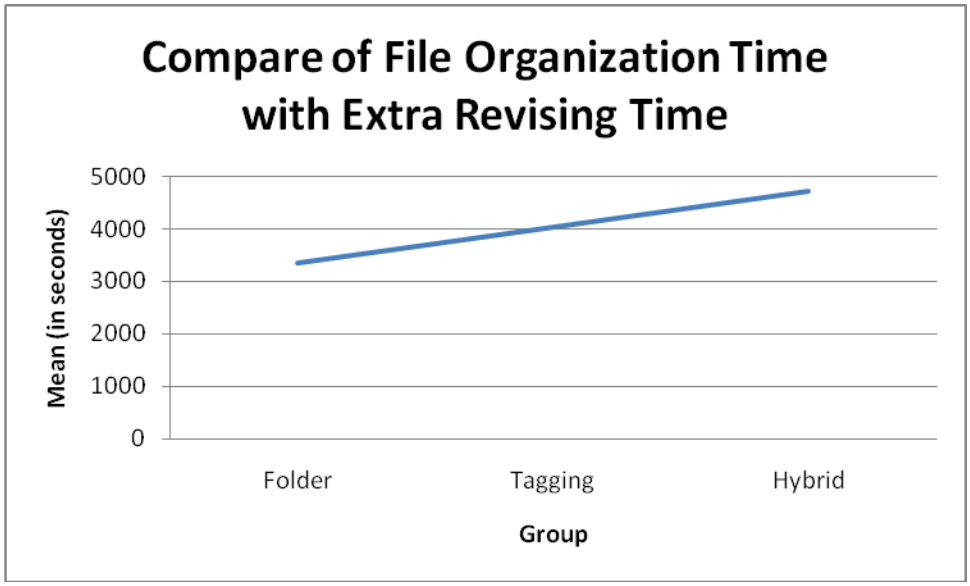


Figure 4. Compare of File Organization Time with Extra Revising Time

4.1.1.2 File Organization Time without Extra Revising Time

A comparison of the file organization time without extra revising time was conducted across the three groups. Table 4 shows the mean and standard deviation of file organization time without extra revising time for the three groups. ANOVA showed that there was a significant difference in the file organization time among the three groups, $F(2, 46) = 10.75$, $p<.001$. Post-hoc Tukey's HSD test showed that the file organization time without extra revising time in the hybrid group was significantly more than the folder group, $p<.001$. Furthermore, the file organization time without extra revising time in the tagging group was significantly more than the folder group, $p<.05$. No significance was found from other pair comparisons. Figure 5 below shows the mean value of the file organization time without extra revising time for the three groups.

Table 4. File Organization Time without Extra Revising Time (in seconds)

	Mean	**Standard Deviation**
The Folder Group	3140.94	938.70
The Tagging Group	4019.00	757.19
The Hybrid Group	4706.18	1159.66

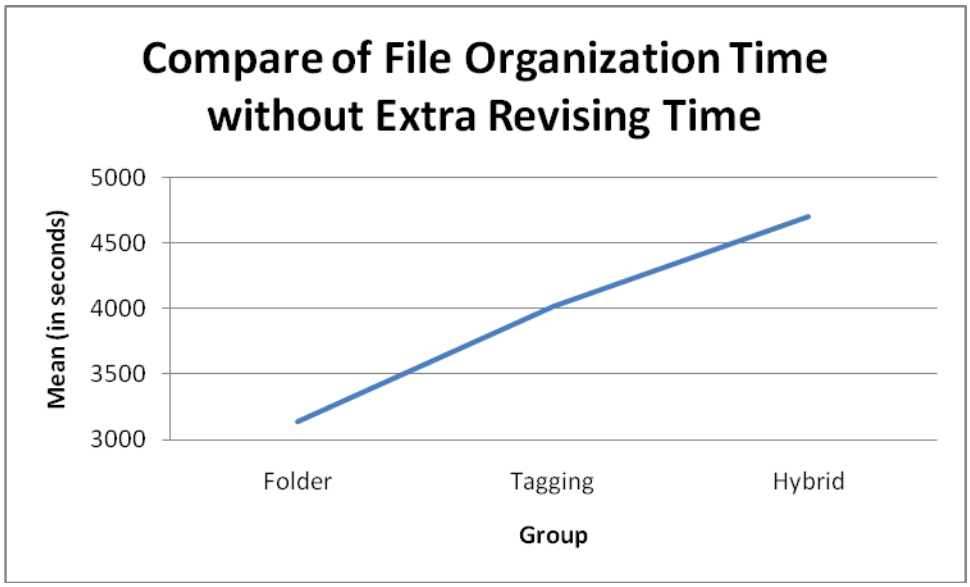


Figure 5. Compare of File Organization Time without Extra Revising Time

4.1.1.3 File Organization Time: Considering the Factor of Unfamiliarity with Tags

One of the concerns in the experimental design is that participants in the tagging group might have spent more time in figuring out how to use tags because they were not familiar with the concept and usage of tags. Although the hybrid and tagging groups had a tutorial on tagging, these groups may not have been fully familiar with the usage of tags, and therefore spent more time in organizing the articles. Participants would have spent more time in the beginning of the information organization figuring out how the system worked.

In order to confirm our prediction, we first compared the time participants in each group spent in organizing the first 30 articles and the second 30 articles to see if there is a significant difference between the time spent in organizing the first 30 articles and the second 30 articles. Table 5 shows the Mean and SD of the organization time spent in the first 30 articles and the second 30 articles for the three groups. ANOVA showed that the folder group did not have a significant difference in the time spent in organizing the first 30 articles and the second 30 articles. For the tagging group, ANOVA showed that participants spent significant more time in organizing the first 30 articles than the second 30 articles, $F(1, 30) = 6.46$, $p < .05$. For the hybrid group, ANOVA showed that participants spent significantly more time in organizing the first 30 articles than the second 30 articles. Since the tagging group and the hybrid group spent less time in the second half of the file organization session than the first half, we believe that the unfamiliarity with tags did play a factor in the time difference observed from the previous comparison.

Table 5. File Organization Time for the First 30 and the Second 30 Files (in seconds)

	First 30 articles		**Second 30 articles**	
	M	**SD**	**M**	**SD**
The Folder Group	1669.69	486.37	1471.25	571.73
The Tagging Group	2193.25	421.40	1808.44	435.21
The Hybrid Group	2592.06	597.78	2114.12	602.59

In order to get a clear understanding of how the level of familiarity with tags affected participants' efficiency in organizing the entire 60 articles, one-way ANOVA was conducted to compared the time spent in organizing every ten files: files 1-10, files 11-20, files 21-30, files 31-40, files 41-50, and files 51-60. The question asked was whether participants spent different organization time in each of the file segments as the organization

session proceeded in each of the three groups. Table 6 shows the Mean and SD of the organization time spent in each of the file segments.

Table 6. File Organization Time for Each File Segment (in seconds)

	Files 1-10		**Files 11-20**		**Files 21-30**	
	M	**SD**	**M**	**SD**	**M**	**SD**
The Folder Group	615.44	179.86	563.00	192.40	491.25	175.10
The Tagging Group	815.56	200.44	741.69	150.11	636.00	138.26
The Hybrid Group	986.35	212.78	828.71	220.82	777.00	209.93

Table 6. (Continued)

	Files 31-40		**Files 41-50**		**Files 51-60**	
	M	**SD**	**M**	**SD**	**M**	**SD**
The Folder Group	470.50	185.56	522.12	229.08	478.62	188.64
The Tagging Group	626.88	169.91	603.06	156.94	578.50	150.60
The Hybrid Group	700.12	175.59	718.65	228.70	695.35	228.50

For the folder group, no significant difference was found between any different file segments. Although we can see the decreasing trend in the organization time from the first file segment to the last file segment, the differences between the segments were not statistically significant.

For the tagging group, we also observed a decreasing trend in the organization time from the first file segment to the last file segment. Post-hoc Tukey's HSD test showed that the organization time in the first file segment (files 1-10) was significantly more than the later 4 file segments (files 21-30, 31-40, 41-50, 51-60), $p<.05$.

For the hybrid group, the decreasing trend in the organization time from the first file segment to the last file segment stayed the same. Post-hoc Tukey's HSD test showed that the organization time in the first file segment (files 1-10) was significantly more than the later 3 file segments (files 31-40, 41-50, 51-60), $p<.05$.

In summary, the overall trend of time spent in file segments was decreasing for all three groups. The first file segment took significantly longer time than the rest file segments only in the tagging group and the hybrid

group. Therefore, we believe that the first file segment (files 1-10) was the only file segment that was affected by participants' unfamiliarity with the tagging structure. Consequently, to account for the unfamiliarity with tags and to eliminate bias, we decided to compare the organization time using article 11 to article 60, instead of the entire article collection.

ANOVA showed that there was a significant difference in the file organization time (from article 11 to article 60) among the three groups, $F(2, 46) = 8.68$, $p<.001$. Table 7 shows the mean and standard deviation of file organization time for the three groups. Post-hoc Tukey's HSD test showed that the file organization time in the hybrid group was significantly more than the folder group, $p<.001$. All other pair comparisons were not significant. The results are consistent with previous comparisons. The figure 6 below shows the mean value of the partial file organization time for the three groups.

Table 7. File Organization Time for Files 11-60 (in seconds)

	Mean	**Standard Deviation**
The Folder Group	2525.50	825.01
The Tagging Group	3186.12	627.33
The Hybrid Group	3719.82	971.40

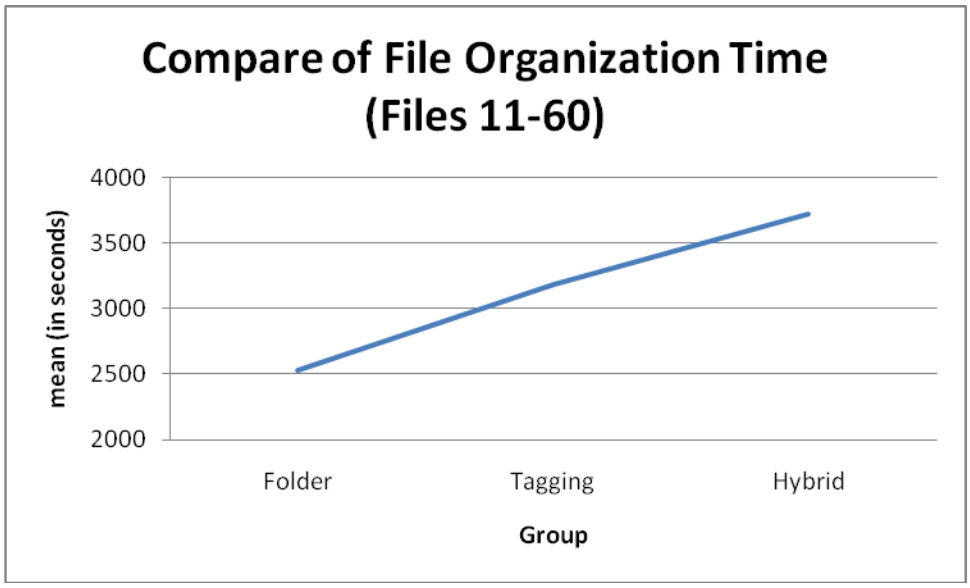


Figure 6. Compare of File Organization Time for Files 11-60

4.1.1.4 Summary and Discussion on File Organization Time

We compared the file organization time with and without extra revising time across the three groups, and compared the partial file organization time (from article 11 to article 60) across the three groups, in order to eliminate the bias from participants' unfamiliarity with tags. Our conclusion was that the hybrid group required longer time than the folder group to organize the same amount of files. No significant difference was found between the folder group and the tagging group or between the hybrid group and the tagging group.

As we expected that the hybrid group would require more time to organize the articles, we were surprised that the difference was only significant between the hybrid and the folder group, but not between the hybrid group and the tagging group. All participants were working on the same corpus of articles in the file organization session. The folder group created various folders and saved the articles in appropriate folders. The tagging group assigned tags to each of the articles. In the hybrid group, participants were required to create both folders and tags. Based on the researcher's observation during the experiment and revisiting of the video recordings after the experiment, the reason for such statistical results seems to be that the hybrid participants spent more effort on the tags than on the folders when given both options. An extreme instance was that one hybrid participant put all 60 articles in one folder, but assigned detailed tags to each of the articles. In other words, if

given limited time and effort in the file organization session, the hybrid group participants were more likely to spend the time in creating tags rather than creating folders.

As for the comparison between the folder structure and the tagging structure, we expected that participants would work differently with the two organization structures, but we were not sure which structure would help participants complete the organize task faster. Indeed, we observed that the folder participants and the tagging participants were quite different in organization behaviors in many aspects. We noticed that seven folder participants spent extra time in revising the self-created folder structures at the end of the session, while none of the tagging participants did the same. As we mentioned previously, we conducted a comparison of file organization time both with and without considering the extra revising time. The results from the two comparisons were different. The tagging and the folder group spent a similar amount of time in file organization if considering the extra revising time. The tagging group spent significantly more time than the folder group in file organization if not considering the extra revising time. Whether or not the extra revising time after the 60 files were saved should be considered as part of the file organization process is an interesting question to ask. The purpose of spending the extra revising time was to revise the folder structure for better re-finding. We think that it should be considered as an integral part of the file organization process, no matter if the revision took place during or after organizing the 60 files.

If we took the factor of unfamiliarity with tags into consideration, the tagging and the folder group spent a similar amount of time in file organization by comparing only time spent from file 11 to file 60. Therefore, there was no difference in the file organization time between the tagging and the folder group. Although participants normally assigned more than one tag to each of the articles, overall it did not take more time than saving the article into one folder. This suggested that assigning a tag is easier than categorizing a file into a folder; however, such an advantage is offset by the fact that usually multiple tags are needed in order to organize a file.

4.1.2 Organization Scheme: Number of Folders and Tags Created

Each participant created a unique organization scheme with folders, tags, or both. The organizational schemes were saved within Zotero automatically. The organization schemes were quantified by counting the numbers of folders, the levels of the folder hierarchy, and the numbers of tags. For the folder group, the number of folders

and the levels of folder hierarchy were counted. For the tagging group, the number of tags was counted. For the hybrid group, the number of folders, the levels of folder hierarchy and the number of tags were counted.

4.1.2.1 Number of Folders

Both participants in the folder group and the hybrid group created folders. The mean value of the folder numbers created by the folder group was 28.75 and standard deviation was 24.65. The mean of the folder numbers created by the hybrid group was 22.18 and the standard deviation was 24.99. ANOVA did not find any significant difference between the two groups.

4.1.2.2 Level of Folders

Both the folder group and the hybrid group participants created multiple levels of folder hierarchies. The mean value of the folder levels created by the folder group was 2.81 and the standard deviation was 0.91. The mean value of the folder levels created by the hybrid group was 1.82 and the standard deviation was 0.81. ANOVA showed that the hybrid group had significantly fewer folder levels than the folder group, $F(1, 31) = 10.91$, $p < .01$. Figure 7 below shows the mean value of folder levels of the two groups.

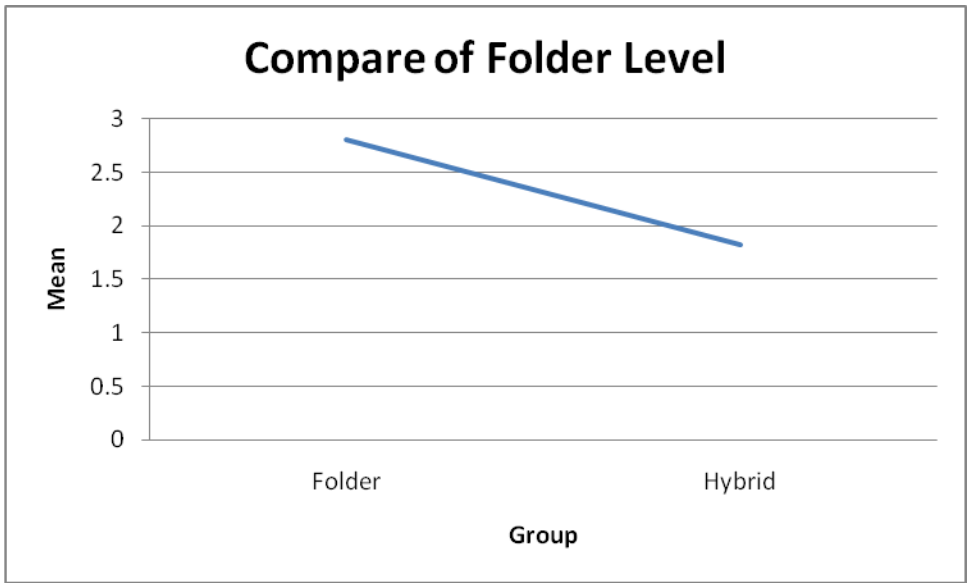


Figure 7. Compare of Folder Level

4.1.2.3 Number of Tags

The comparison of the number of tags created was conducted between the tagging group and the hybrid group. The mean number of tags created by the tagging group was 97.50, and standard deviation was 61.06. The mean number of tags created by the hybrid group was 147.76 and the standard deviation was 73.53. ANOVA showed that the tagging group created significantly less number of tags than the hybrid group, $F(1, 31) = 4.53$, $p < .05$. Figure 8 below shows the mean value of tag number created by the tagging group and the hybrid group.

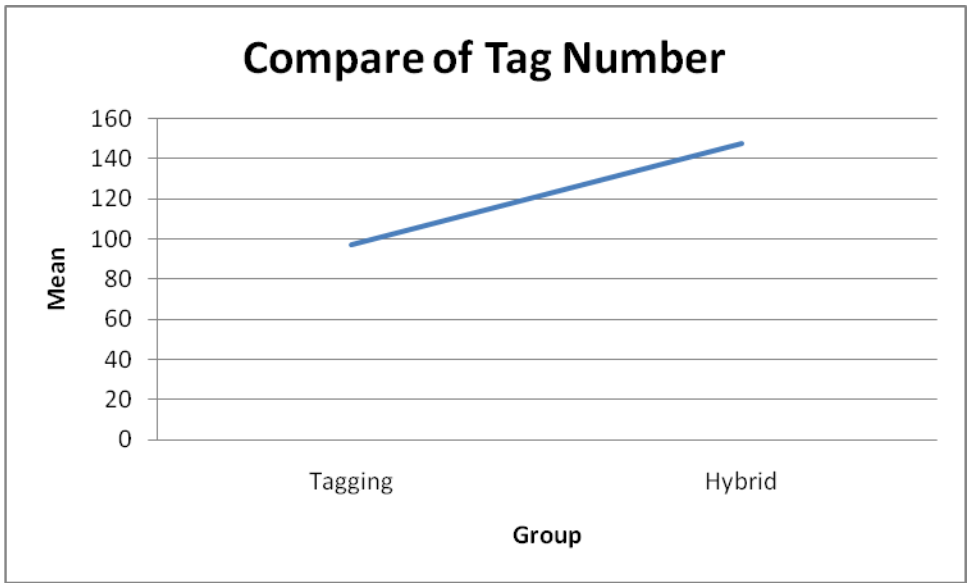


Figure 8. Compare of Tag Number

4.1.2.4 Discussion on File Organization Scheme

The comparison of number of folders and folder level was between the folder and the hybrid group, as the tagging group did not create any folders. The results showed that the two groups were not significantly different in the number of folders, but significantly different in the levels of folders. The hybrid group created as many folders as the folder group, but the structure was not as deep as the folder group. The comparison of number of tags was conducted between the tagging and the hybrid group, as the folder group did not create any tags. The results showed that the hybrid group created significantly more tags than the tagging group.

The results confirmed our observation mentioned previously that the hybrid group participants tended to spend more effort in creating tags than creating folders when given both options. Between the folder and the hybrid group, the number of folders stayed the same. The fewer levels of folders indicated less effort in creating hierarchies. The hybrid participants were satisfied with simply putting the article into a folder, without much effort thinking about how folders could be combined into hierarchies. However, they created more tags to make up the weak folder structures, in order for them to be able to re-find files in the later session.

4.2 How Were the Three Groups Different in the File Re-finding Session?

4.2.1 File Re-finding Time

The file re-finding session included four tasks. For each task, the task completion time was measured from the time participants started to read the task description for the current task until they finished filling answers in the Answer Sheet. The time information was captured by the screen recordings. In order to find out how much time participants spent in completing each task and the entire file re-finding session, the researcher watched the video recording closely and marked the starting and ending point for each task. Comparisons of the time spent in completing each task in the file re-finding session were conducted across the three groups. The means and standard deviations of the time spent in each of the re-finding tasks are listed in table 8.

Table 8. Compare of File Re-finding Time (in seconds)

	Task 1		**Task 2**		**Task 3**		**Task 4**	
	M	**SD**	**M**	**SD**	**M**	**SD**	**M**	**SD**
The Folder Group	130.44	37.94	215.94	65.85	297.62	132.51	359.62	136.12
The Tagging Group	133.31	62.36	214.50	102.75	294.56	125.33	364.62	148.76
The Hybrid Group	142.65	57.03	243.18	61.92	277.76	116.40	378.59	172.40

In re-finding task 1, participants in all three groups were given five file titles as clues, based on which they were asked to re-find the target article. ANOVA showed no significant difference in the re-finding time for task 1.

In re-finding task 2, participants in all three groups were given five article summaries as clues, based on which they were asked to re-find the target article. ANOVA showed no significant difference in the re-finding time for task 2.

In re-finding task 3, participants in all three groups were given five questions. Each question provided an information item that was mentioned in one of the 60 articles. Based on the question, participants were asked to re-find the target file. ANOVA showed no significant difference in the re-finding time for task 3.

Finally, in re-finding task 4, participants in all three groups were given three questions. Each question asked an information item that was relevant to 6 of the 60 files. Based on the question, participants were asked to re-find all the 6 files. ANOVA showed no significant difference in the re-finding time for task 4.

In summary, the analysis results showed that there was no difference among the re-finding time across the three groups.

We expected to see a difference in the re-finding times across the three groups, however, no significant difference was found. A possible reason for the results was that some participants might simply have left the answer box empty if they did not want to spend too much time on it. Therefore, we compared the time spent in only re-finding the correct articles. However, this effort did not yield any significant results either. In conclusion, the three groups performed equally in terms of how fast they finished the re-finding tasks.

4.2.2 Answer Correctness

Participants' answers were saved in Word documents. The answers were checked against the researcher's correct answers to determine how many files were correctly found, how many files were wrongly found, and how many files were left unfound. We developed an answer correctness score for the correct answer, wrong answer and skipped answer. Participants gain one score by finding a correct file; they lose one score by finding a wrong file; they do not gain or lose by skipping a file.

For re-finding task 1, task 2 and task 3, the highest score was 5, in which case the participant had correctly found all the five target files. The lowest score was 0 in which case the participant had not found any of the target articles. Re-finding task 4 had three subtasks that each requiring finding multiple articles. The highest score for each subtask was 6, in which case the participant had found all the 6 target files. The lowest score was 0 in which case the participant had not found any of the target files. The means and standard deviations of the answer correctness for each of the re-finding tasks are in table 9.

Table 9. Answer Correctness for the File Re-finding Tasks

	Task 1		Task 2		Task 3		Task 4	
	M	**SD**	**M**	**SD**	**M**	**SD**	**M**	**SD**
The Folder Group	4.88	0.50	4.69	0.48	4.00	0.82	3.85	1.35
The Tagging Group	4.88	0.50	4.25	1.13	3.44	1.15	3.75	1.62
The Hybrid Group	4.71	0.58	4.53	0.80	3.41	0.80	3.59	1.36

ANOVA showed no statistical significance in the answer correctness across three groups in all re-finding tasks. The three groups performed equally in terms of how correctly they re-found files.

4.2.3 Number of Mouse Clicks

The number of mouse clicks was counted for the file re-finding session. A mouse click is defined as a click on a folder or a tag. The number of mouse clicks was counted manually by the researcher based on the screen recordings. For each re-finding task, the number of mouse clicks was compared across the three groups. For the hybrid group, we also conducted a comparison between the number of folder clicks and the number of tag clicks.

4.2.3.1 Number of Mouse Clicks in Task 1

In re-finding task 1, participants in all three groups were given five article titles as clues, based on which they were asked to re-find the target file. The number of mouse clicks was counted for each subtask. A comparison of the number of mouse clicks was conducted across the three groups. Table 10 shows the means and standard deviations of mouse click numbers for the three groups. ANOVA showed that there was a significant difference in the number of mouse clicks among the three groups, $F(2, 234) = 4.66$, $p < .01$. Post-hoc Tukey's HSD test showed that the folder group used significantly more numbers of mouse clicks than the tagging group on task 1, $p<.05$. All other pair comparisons were not significant. Figure 9 below shows the mean value of task 1 mouse click numbers for the three groups.

Table 10. Task 1 Mouse Click Number

	Mean	**Standard Deviation**
The Folder Group	2.57	3.53
The Tagging Group	1.41	1.49
The Hybrid Group	1.86	1.61

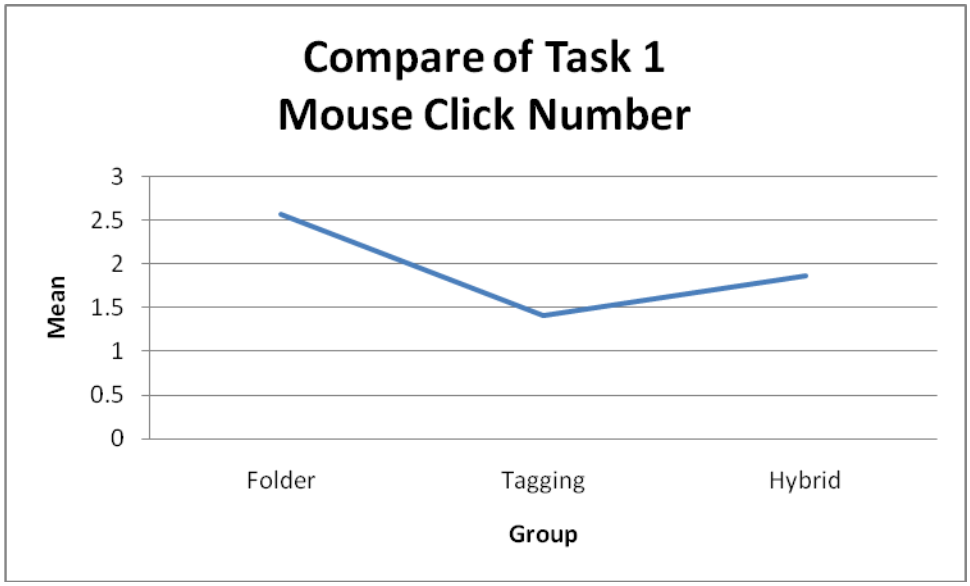


Figure 9. Compare of Task 1 Mouse Click Number

4.2.3.2 Number of Mouse Clicks in Task 2

In task 2, participants in all three groups were given five article summaries as clues, based on which they were asked to re-find the target article. Similarly, the number of mouse clicks was counted for each subtask. A comparison of the number of mouse clicks was conducted across the three groups. Table 11 shows the means

and standard deviations of the numbers of mouse clicks for the three groups. ANOVA showed that there was a significant difference in the number of mouse clicks among the three groups, $F(2, 221) = 3.58$, $p<.05$. Post-hoc Tukey's HSD test showed that the folder group used significantly more numbers of mouse clicks than the tagging group on task 2, $p<.05$. All other pair comparisons were not significant. Figure 10 below shows the mean of task 2 mouse click numbers for all the three groups.

Table 11. Task 2 Mouse Click Number

	Mean	**Standard Deviation**
The Folder Group	2.78	5.55
The Tagging Group	1.39	0.84
The Hybrid Group	2.66	2.11

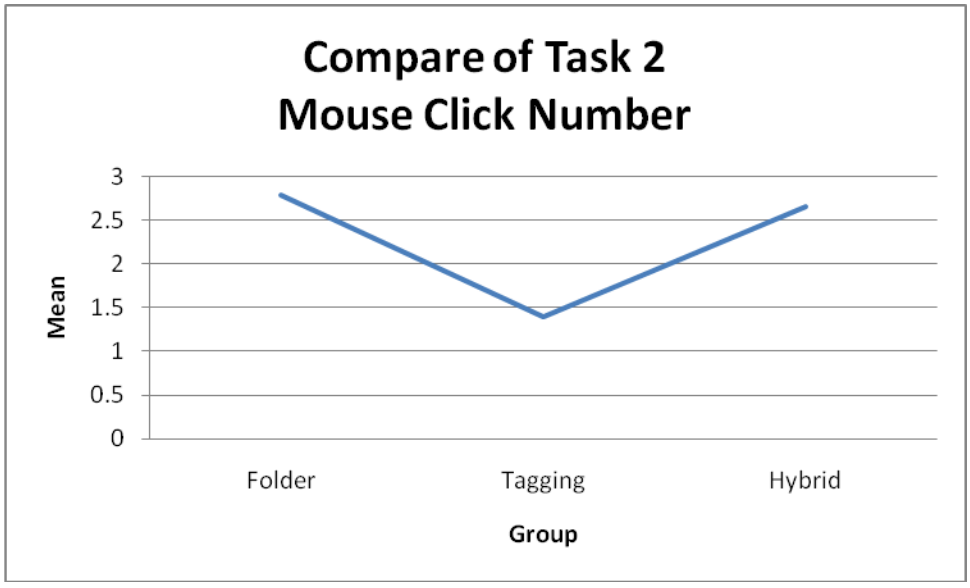


Figure 10. Compare of Task 2 Mouse Click Number

4.2.3.3 Number of Mouse Clicks in Task 3

In task 3, each question provided a piece of information that mentioned the content of one of the 60 articles. Based on the question, participants were asked to re-find the target article. Similar to task 1 and task 2, the number of mouse clicks was counted for each subtask. A comparison of the number of mouse clicks was conducted across the three groups. Table 12 shows the means and standard deviations of number of mouse clicks for the three groups. ANOVA showed that there was a marginal difference between the groups in the number of mouse clicks, $F(2, 186) = 2.55$, $p < .08$. Similar to the previous two tasks, the tagging group used a fewer number of mouse clicks than the folder group. Figure 11 below shows the mean of task 3 mouse click numbers for the three groups. The figure shows a pattern similar to the previous two tasks.

Table 12. Task 3 Mouse Click Number

	Mean	Standard Deviation
The Folder Group	3.45	6.17
The Tagging Group	1.78	1.46
The Hybrid Group	2.75	3.01

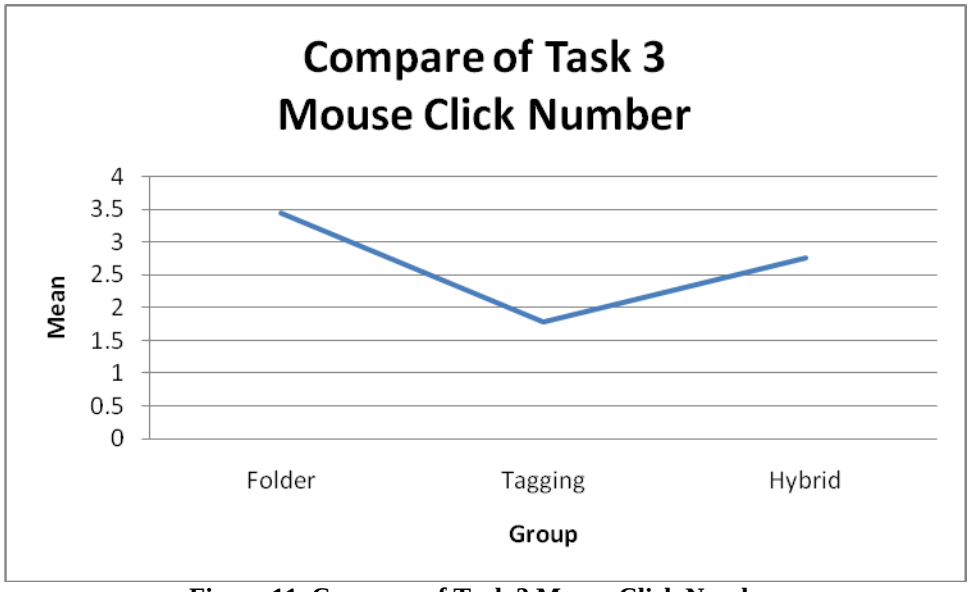


Figure 11. Compare of Task 3 Mouse Click Number

4.2.3.4 Number of Mouse Clicks in Task 4

In task 4, participants in all three groups were given three questions. Each question asked a piece of information that was relevant to 6 of the 60 articles. Based on the question, participants were asked to re-find all 6 articles. Similarly, number of mouse clicks was counted for each subtask. A comparison of the number of mouse clicks was conducted across the three groups. Table 13 shows the means and standard deviations of mouse click numbers for the three groups. ANOVA showed that there was a significant difference in the number of mouse clicks across the three groups, $F(2, 143) = 3.74$, $p < .05$. Post-hoc Tukey's HSD test showed that the hybrid group used significantly more numbers of mouse clicks than the tagging group, $p < .05$. All other comparisons were not significant. Figure 12 below shows the mean of task 4 mouse click numbers for the three groups.

Table 13. Task 4 Mouse Click Number

	Mean	Standard Deviation
The Folder Group	8.79	5.49
The Tagging Group	6.43	4.21
The Hybrid Group	9.35	6.66

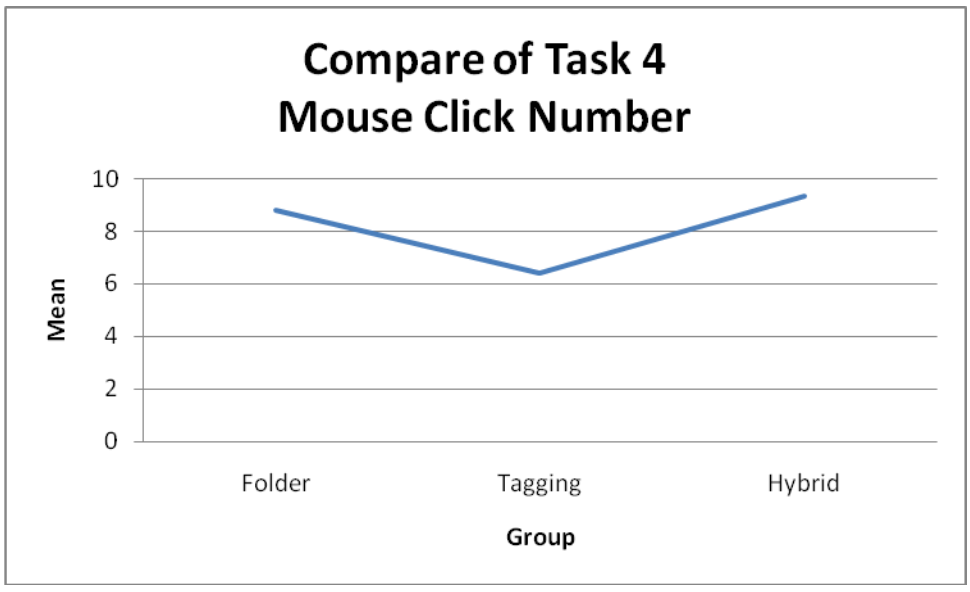


Figure 12. Compare of Task 4 Mouse Click Number

4.2.3.5 Hybrid Group Folder Clicks and Tag Clicks

Participants in the hybrid group could click on both folders and tags to re-find the target files. We compared the number of folder clicks and the number of tag clicks for each re-finding task. Table 14 shows the means and standard deviations of mouse clicks for all the tasks. Participants used more folder clicks than tag clicks in re-finding task 1, task 2 and task 3. They used similar number of mouse clicks in task 4. ANOVA showed that the difference between folder clicks and tag clicks was significant in task 2, $F(1, 156) = 4.29$, $p < .05$.

Table 14. Hybrid Group Folder Clicks and Tag Clicks

	Task 1		**Task 2**		**Task 3**		**Task 4**	
	M	**SD**	**M**	**SD**	**M**	**SD**	**M**	**SD**
Folder Clicks	1.01	1.47	1.62	1.95	1.75	3.05	4.67	5.72
Tag Clicks	0.84	0.86	1.06	1.38	1.00	1.16	4.69	5.64

4.2.3.6 Discussion on File Re-finding Mouse Clicks

The number of mouse clicks was used as the measurement for how much cognitive effort was used in completing the re-finding tasks. The more number of mouse clicks required in a re-finding task, the higher level of cognitive effort used in the re-finding process, and therefore the more frustration experienced by the user. For the folder group, a mouse click was defined as a click on a folder to view files inside that particular folder. For the tagging group, a mouse clicks was defined as a click on a tag to view the files that were associated with the particular tag. For the hybrid group, a mouse click could be either on a folder or on a tag. The number of mouse clicks in the hybrid group was a summary of clicks on folders and tags.

We designed four different re-finding tasks in order to simulate a real working environment. Users remember different things about the target files that they want to re-find. The simplest case would be that they remember exactly what the article title was, which was the situation of task 1. Task 2 was the situation where users only remember the overall content of the article, which we simulated in the form of a content summary. In the case

where users forget the exact title as well as the overall content, what they do remember is a specific item of information mentioned in the article. In task 3, we used a question that provided an item of information that was mentioned in one of the articles. There were also situations where users needed to find multiple relevant articles to answer a question. In task 4, each question asked an item of information that was relevant to six of the 60 articles and participants had to find all of them.

The comparisons of mouse clicks in all of the four re-finding tasks showed that the three groups were significantly different. For task 1, task 2 and task 3, the tagging group used the fewest number of mouse clicks, the folder group used the most number of mouse clicks and the hybrid group fell in between. A significant difference existed only between the tagging and the folder group. For task 4, the tagging group used the fewest number of mouse clicks, the hybrid group used the most number of mouse clicks and the folder group fell in between. The significant difference existed only between the tagging and the hybrid group. In all four re-finding tasks, by using the fewest number of mouse clicks, we believe that the tagging group participants experienced less cognitive effort in re-finding than the other two groups.

The tagging group used fewer mouse clicks than the folder group in all four re-finding tasks. As we discussed in Chapter 2, the advantage of using tags for re-finding is that a file can be accessed with any associated tags, while the folder structure required users to follow a linear path in order to re-find the files. Usually an article was associated with multiple tags and any of these tags could be used for re-finding. Tagging helped participants to reach the target articles with fewer mouse clicks. For instance, if an article was assigned with five tags, the tagging group participant could re-find the article by clicking on any of the five tags. In the folder group, the same article was put into a folder. In order to re-find the article, the folder group participant would have to click on this particular folder to see the target article. In the ideal situation, when participants remembered perfectly all the associated tags or the exact folder that the file was saved, all participants would be able to access the file with one single click either on one correct tag or the correct folder. However, in a real situation, when participants usually did not remember everything perfectly, in most cases they had to click on multiple folders or tags in order to navigate to the correct one. A larger number of access points (tags associated with the target file) in the tagging group made it more likely for the file to be accessed with fewer clicks.

However, this does not mean that the more tags an article was assigned to, the easier the article was re-found. When too many tags were assigned to an article, many tags were general words. Such tags did not help participants to pinpoint the target article. Also, increasing tag numbers could be associated with other problems, such as tagging navigation inefficiency. Participants had to scroll up and down the tag list many times to locate a particular tag.

Initially, we expected that the hybrid structure would be more helpful than the folder and the tagging structure in re-finding the articles, as it provided support for both folders and tags. Since we did not see any advantages from the hybrid group in previous comparisons of re-finding time and answer correctness, we suspected that the hybrid group would use the fewest number of mouse clicks in the re-finding tasks. However, we found that instead of the hybrid group, the tagging group used the fewest number of mouse clicks. The researcher went back to the screen capture video for more detailed observation. The results will be discussed in Chapter 5, which will qualitatively focus on the mouse clicks in the re-finding process.

We saw different patterns in the number of mouse clicks in the first three tasks and task 4. In task 1, 2, and 3, the hybrid group fell between the tagging and the folder group. In task 4, the folder group fell between the tagging and the hybrid group. The comparison of folder clicks and tag clicks in the hybrid group revealed that the hybrid group participants had a higher number of folder clicks than tag clicks in task 1, task 2 and task 3. In task 4, though, the number of tag clicks was more than folder clicks. The primary difference between task 1, task 2, task 3 and task 4 was that the first three tasks asked the participants to re-find single articles, while task 4 asked them to re-find multiple articles. It seems that re-finding multiple files created a very different working environment from re-finding single files. We will discuss the interplay between re-finding task difference and the organization structures in Chapter 5.

4.3 How Were the Three Groups Different in User Perception?

4.3.1 User Perception in the File Organization Session

At the end of the file organization session, participants filled out a questionnaire that asked about their perception of their own organization performance. Participants were asked to rate the following items: perceived successfulness in carrying out the organization task, how easy it was to use Zotero, how easy it was to

understand the articles, how easy it was to organize the articles, how easy it was to create folders or tags, and overall satisfaction with the self created organization structure (See appendix C for questionnaire details). For each of these items, participants gave a rating between 1 and 5. Table 15 shows the means and standard deviations of the perception items in the file organization session. ANOVA was used to compare the variables across the three groups. No significant difference was found among the three groups in all the above perceived items.

Table 15. User Perception in the File Organization Session

	Perceived successfulness		**Ease of use**		**Ease of understanding the articles**		**Ease of organizing the articles**		**Ease of creating folders/tags**		**Overall satisfaction**	
	M	**SD**	**M**	**SD**	**M**	**M**	**SD**	**SD**	**M**	**SD**	**M**	**SD**
The Folder Group	4.69	0.48	4.56	0.51	4.56	0.51	3.88	0.50	3.56	0.73	4.25	0.45
The Tagging Group	4.69	0.48	4.44	0.63	4.69	0.48	4.06	0.44	4.06	0.57	3.88	0.72
The Hybrid Group	4.76	0.44	4.47	0.51	4.47	0.62	3.82	0.53	3.65 4.29	0.70 0.59	3.94	0.66

Since the hybrid group had rated their perceived ease of creating both folders and tags, we conducted a comparison between the ease of creating folders and ease of creating tags within the hybrid group. ANOVA showed that the hybrid group participants believed that it was significantly easier to create tags than to create folders, $F(1,32) = 8.49$, $p < .01$. Figure 13 shows the mean value of the perceived ease of creating folders and creating tags within the hybrid group.

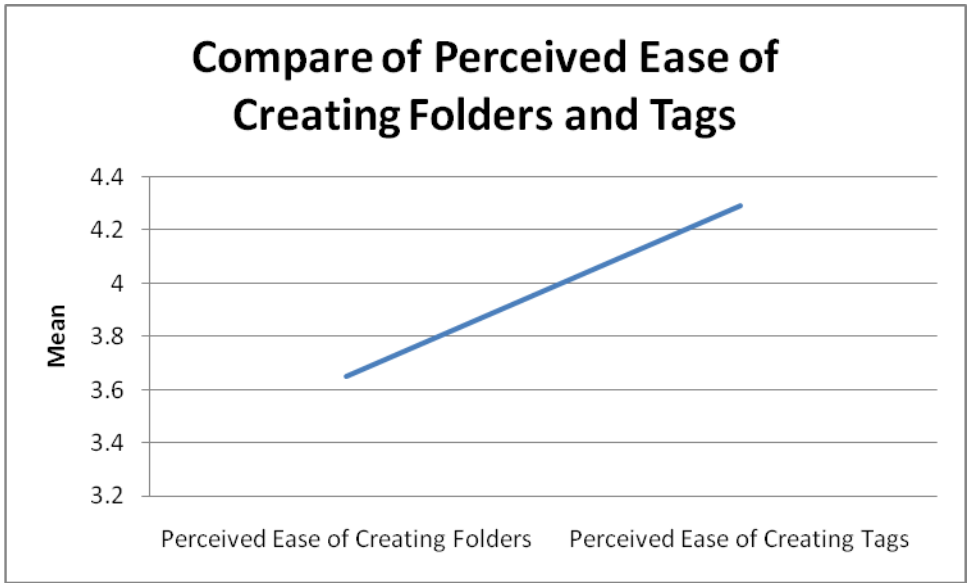


Figure 13. Compare of Perceived Ease of Creating Folders and Tags in the Hybrid Group

4.3.2 User Perception in the File Re-finding Session

At the end of the file re-finding session, participants filled out a questionnaire that asked about their perception of their own re-finding performance. Participants were asked to rate the following items: perceived successfulness in carrying out the re-findings tasks, how easy it was to re-find files, how helpful the folders were in re-finding, how useful the folders were in providing reminding functions, how helpful the tags were in re-finding, and how useful the tags were in providing reminding functions. For each of these items, participants gave a rating between 1 and 5. Table 16 shows the means and standard deviations of the perception items in the file re-finding session.

Table 16. User Perception in the File Re-finding Session

	Perceived successfulness		Ease of re-finding		Helpfulness of folders/tags		Reminding function of folders/tags	
	M	**SD**	**M**	**SD**	**M**	**SD**	**M**	**SD**
The Folder Group	4.56	0.51	3.94	0.57	4.56	0.63	4.44	0.89
The Tagging Group	4.19	0.66	4.12	0.50	4.31	0.70	4.44	0.51
The Hybrid Group	4.12	0.70	3.94	0.43	3.88	0.93	4.00	1.00
					4.41	1.23	4.35	1.12

ANOVA found that the perceived helpfulness of folders in the hybrid group was significantly lower than the perceived helpfulness of folders in the folder group, $F(1, 31) = 5.99$, $p < .05$. Figure 14 shows the mean of the perceived helpfulness of folder in the hybrid group and the folder group.

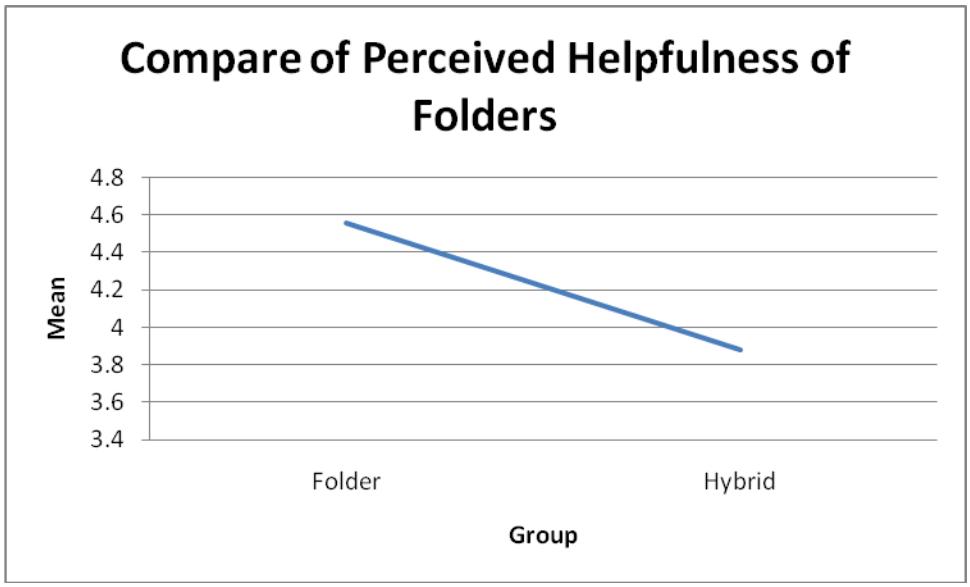


Figure 14. Compare of Perceived Helpfulness of Folders

Since the hybrid group had rated their perceived helpfulness and the reminding function of both folders and tags, we conducted a comparison between folders and tags within the hybrid group. Although the ratings for tags were higher than folders in both of the two variables, ANOVA showed no significant difference between them for the hybrid group participants.

In conclusion, the participants in all three groups had similar levels of perceived successfulness and perceived ease of carrying out the file organization tasks and file re-finding tasks. There were some differences in terms of creating folders and tags, and using folders and tags for re-finding tasks. The hybrid participants believed it was easier to create tags than to create folders, which supported our discussion in the file organization session.

Chapter 5: Qualitative Discussion on File Organization and File Re-finding

This chapter will be focused on the qualitative findings from the file organization session and the file re-finding session. We will first discuss the organization behaviors observed in the file organization session, and then discuss the re-finding strategies observed in the file re-finding session.

5.1 Organizing Files with Folders and Tags

In the first experimental session, participants were given the same 60 articles but working with three different organization structures. Consequently they created very different organizational schemes. This section discusses how folders and tags were created by using different organization structures.

5.1.1 Previous Models on Tagging and Categorizing

Pak et al. (2007) conducted two studies to examine the differences between hierarchical categories (folders) and tags for text organization and retrieval. They proposed a cognitive model for categorization (creating folders) and another model for tagging, by revising Sinha's model (2005). Pak et al.'s categorization model depicted that multiple "concepts" were activated after the to-be-organized information was understood and processed, and that an appropriate category would either be created (if no appropriate folders existed) or selected (if an appropriate folder already existed). For the tagging process, similarly, multiple "concepts" were activated after the to-be-organized information was understood and processed. Then tags were created for these concepts as descriptive labels. Our experiment did not only shed lights on the cognitive process of how folders and tags were created individually, but also uncovered the cognitive process users went through when they needed to create both folders and tags for the same to-be-organized information.

5.1.2 Different Organization Strategies in Creating Folders and Tags

Malone (1983) proposed two different strategies for desktop file management, filers and pilers. Filers refer to people who maintained precisely characterized piles. Pilers refer to people whose offices were filled with miscellaneous piles. Whittaker (1996) identified three strategies for handling email overload: *no filers* (no use of folders), *frequent filers* (folder users who try to clean up their inbox daily) and *spring cleaners* (folder users

who clean up their inbox only periodically). In our experiment, we also observed different strategies in organizing articles.

5.1.2.1 The Folder Group Organization Strategy

For the folder group, we went through the 17 folder structures created by the folder group and used a card sorting technique to categorize them into different kinds of organizing behaviors (16 participants completed both the file organization and the file re-finding session. One participant only completed the file organization session. The organization scheme created by this particular participant was also used for this analysis). Three different kinds of folder patterns emerged: *Regular Categorizers*, *Extensive Categorizers*, and *Careful Categorizers*. Nine participants created less than 20 folders for the 60 articles. These folder names were general enough that at least three files resided in one folder. Since a majority of the folder participants fell in this group, we call them *Regular Categorizers*. Four participants created 20 to 60 folders for the 60 articles, whom we call *Extensive Categorizers*. These participants usually created more specific subfolders under general parent folders, for example, a parent folder was created as "technology", a subfolder being "computer" and a sub-subfolder being "hardware". The rest four participants were called *Careful Categorizers*, because they created more than 60 folders for the 60 articles. The reason that the number of folders was higher than the number of files was that some parent folders only had subfolders in them and did not contain any files, for example, a parent folder was created with the folder name of "gadgets" and a subfolder was created with the folder name of "$5,000 3D Paper Dispenser". The participant put an article about a $5,000 3D paper dispenser into the subfolder of "$5,000 3D Paper Dispenser" and no article was put in the parent folder of "Gadgets." For the subset of participants who created more numbers of folders than the number of files, there was only one file stored in a folder or subfolder in most cases.

In order to learn how organization schemes evolved over time, we drew a chart that showed the accumulation of folders over time. *Careful Categorizers* constantly created new folders as they organize new files in the organization session, until they finished reading the 60 articles. We see a linear growth pattern of the folder number over time. *Extensive Categorizers* created detailed subfolders and maintained a flatter growth path. *Regular Categorizers* created less than 20 folders for the 60 articles and the growth path was even flatter than the *Extensive Categorizers'* growth path.

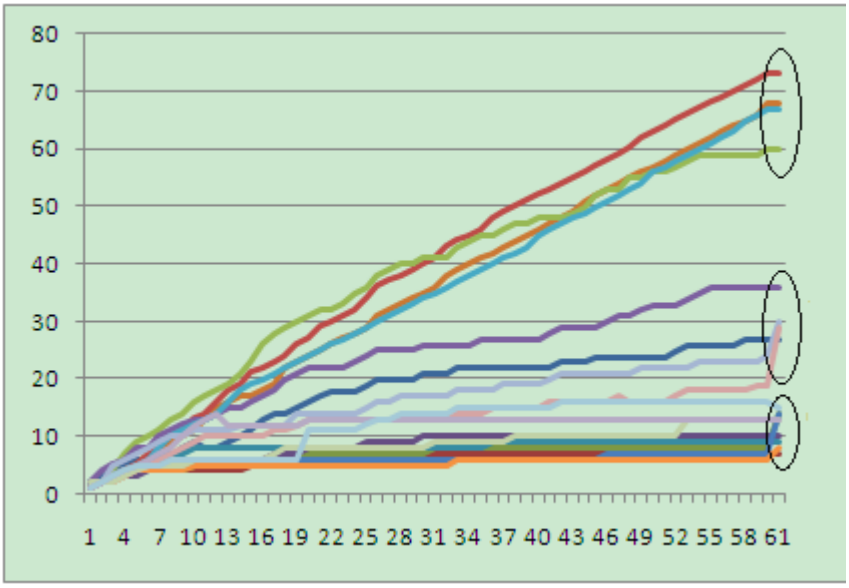


Figure 15. The Growth of Folder Number over Time in the Folder Group

5.1.2.2 The Tagging Group Organization Strategy

For the tagging group, similarly, we analyzed the organization schemes created by the 16 participants who completed both the file organization and file re-finding session, and the one participant who only completed the file organization session. The 17 tagging structures created by the tagging group were grouped into three different kinds of tagging patterns by using a card sorting technique: *Conservative Taggers*, *Regular Taggers*, and *Careful Taggers*. Two participants created less than 20 tags for the 60 articles, who we call *Conservative Taggers*. The tags they created were highly generalized and they were reused intensively, such as "personal computing," "future technology". Usually one tag was associated with multiple files. Four participants were called *Careful Taggers* because they created more than 150 tags for the 60 articles. Eleven participants in the tagging group created a total of less than 150 tags for the 60 articles**.** Since a majority of tagging participants fell into this group, we call them *Regular taggers.* Both *Careful Taggers* and *Regular Taggers* maintained a linear growth path. *Conservative Taggers*, on the other hand, had a flat growth path.

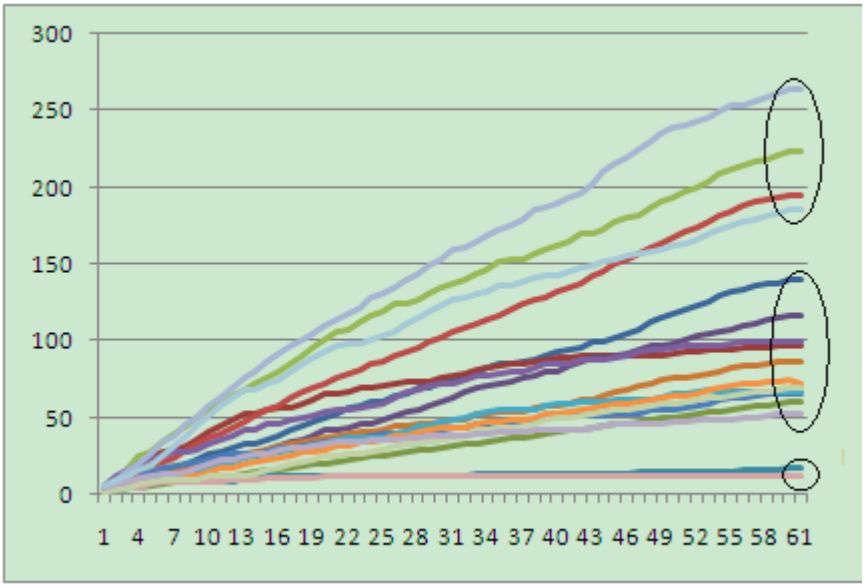


Figure 16. The Growth of Tag Number over Time in the Tagging Group

5.1.2.3 The Hybrid Group Organization Strategy

We also looked at how the numbers of folders and tags growed in the hybrid group. Separate charts on folder growth and tag growth were created. For the growth of folder number over time, we observed similar patterns as the folder group. As indicated by the chart, we saw *Careful Categorizers*, *Extensive Categorizers* and *Regular Categorizers*. For the growth of tag number, we saw *Careful Taggers* and *Regular Taggers*, but missed *Conservative Taggers*.

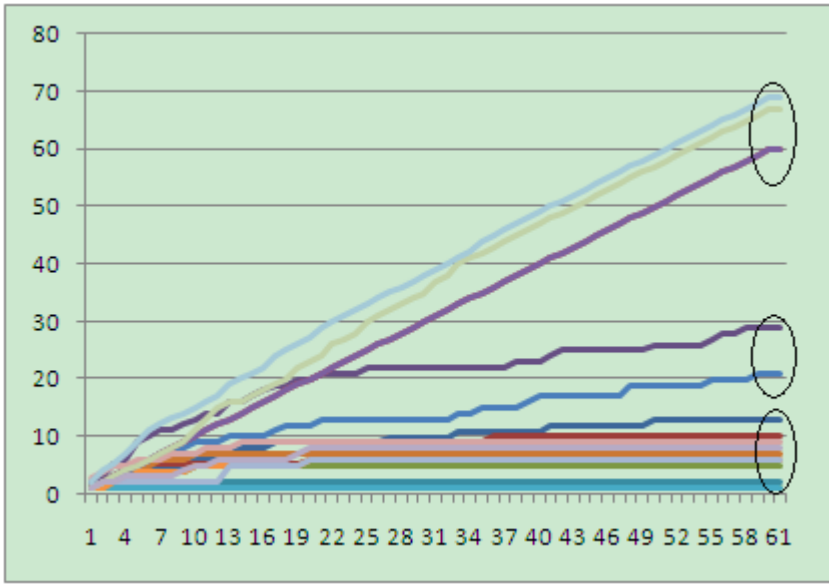


Figure 17. The Growth of Folder Number over Time in the Hybrid Group

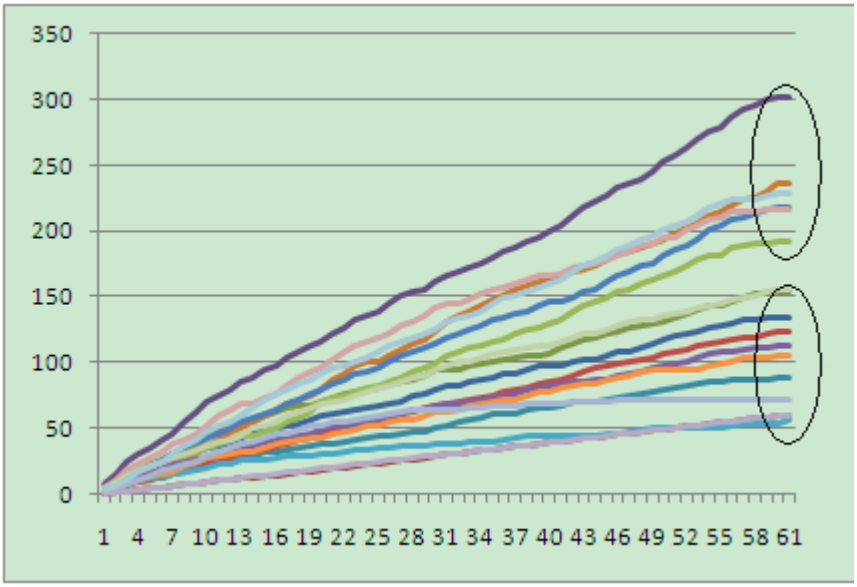


Figure 18. The Growth of Tag Number over Time in the Hybrid Group

5.1.3 Single-word and Multi-word in Folder Names and Tag Names

For each group, a list of the folders or tags created by the group was collected and duplicated ones were eliminated. The number of single-word folder and tag names and multi-words folder and tag names were counted. Of the 417 folders created by the folder group, 24.7% were single word names and 75.3% contained multiple words. Of the 342 folders created by the hybrid group, 28.7% were single-word folder names and 71.3% contained multiple words. Of the 892 tags created by the tagging group, 59.4% were single-word tags and 40.6% were multi-word tags. Of the 1057 tags created by the hybrid group, 61.8% were single word tags and 38.2% were multi-word tags.

Overall, there were more multi-words folder names than single-word folder names, in both the folder group and the hybrid group. Folder names are more likely to be short phrases than single words. There were more single-word tag names than multi-word tag names, in both the tagging group and the hybrid group. Tags are more likely to be single words than short phrases. We believe that the reason for such difference is that folder names were created to provide a content description of all the articles within the folder. It usually requires phrases in order to convey a complete meaning, whereas tags were created to describe one specific information item from one article, in which case single words can usually serve the needs.

5.1.4 Folder Names, Tag Names and the Original Article Text

For the folder group, the list of folders created by the group was compared against the original article text. The folder group created 417 unique folders in total. Among these folders, 35.7% folder names were identical to certain parts of the article text, while 63.5% folder names were different from the original article text. They were generated by rephrasing, paraphrasing or generalizing from the article text.

The list of folders created by the hybrid group was also compared against the original article text. The hybrid group created 336 unique folders in total. Among them, 42.3% folder names were identical to certain parts of the article text, and 55.4% folder names were generated by rephrasing, paraphrasing, or generalizing from the article text.

Using a similar method, the list of tags created by the tagging group was compared against the original article text. The tagging group created 892 unique tags. Of these tags, 67.9% were identical to certain parts of the

original article text, and 30.6% tags were generated by rephrasing, paraphrasing or generalizing from the article text.

The hybrid group created 1057 unique tags. 73.8% tags were identical to certain parts of the original articles text. 33.0% tags were generated by rephrasing, paraphrasing or generalizing from the article text.

Overall, in both the folder and the hybrid group, folder names were more likely to be created by rephrasing, paraphrasing, or generalization from the article text than by directly using the article text. In both the tagging and the hybrid group, tagging names were more likely to be created by directly using the article text than by rephrasing, paraphrasing or generalization. We believe that the reason for such difference was that the creating folder names required a higher level of information procession than creating tags. As Sinha (2005) and Pak et al. (2007) discussed in their research, tags could be any concepts activated by reading the article. Therefore, they were more likely to be directly words chosen from the article text. The creation of folders needed generalization and categorization. Therefore, they were more likely to be different from the original article text.

5.1.5 Participants' Feedback from the Questionnaires

The folder participants reported how they created folders and tags in the file organization session. In terms of how they put multiple articles into one folder, they reported that they tried to find similarities across multiple articles. They attempted to create folder names that were general enough to accommodate several articles. Some participants mentioned that it was necessary to make revisions at the end of the organization session because they had a better overall picture of the 60 article at the end of the organization session to create a better generalization.

Tagging participants reported that they chose the tags that would be easy to recall. Participants created both general tags, such as overall themes that were applied to multiple articles, for example, "internet"; "technology", as well as specific tags, such as author name, brand name, and abbreviations that were only applied to single articles, for example, "LG Philips", "nike +".

Participants in each group reported problems that they experienced in the file organization session. In the folder group, participants reported that articles had overlapping content and they had to spend extra time determining the most important topic before choosing an appropriate folder. A few previously created folders were no longer

appropriate when they read more articles. Participants made changes to folder names and folder structure revisions during the organization process. Several participants reported that some articles could fit into multiple categories but some others did not fit into any, such situation made the organization difficult. Consequently, some folders contained many articles, while some contained very few. Some participants had to create an "others" folder to accommodate those files that can not fit into any folders. Participants also reported the struggle of "finding a perfect folder" for certain files. They reported that it was very difficult to place the article into a category that was specific enough for the file content, "I knew some articles should be further organized, but I could not find a way to."

The tagging group also reported problems when organizing files with tags. Similarly, participants' ideas of appropriate tags changed as they read more articles. For example, one participant mentioned, "sometimes I realized a link [common theme] between several articles after having read a few of them. This meant I had to go back and search for the articles and add the new tag that I thought of." Since most participants were new to the tagging structure, some of them were not sure whether the tags they assigned were going to work in the re-finding session.

The hybrid participants reported their problems in the file organization experience. Participants expressed their struggles with the organization task. "There were a few articles I was not sure where to place them. Some could fit into multiple categories; some were hard to find keywords [tags]." The hybrid participants had problems from creating both folders and tags.

5.1.6 Conclusion on File Organization Observation

File organization is a personal behavior, and individual differences were obvious. In our experiment, we observed a range of differences in the organization behaviors between and within the three groups. However, overall we observed similar patterns within each organization structure. More multi-words were used as folder names than singe words. More folder names were created by generalization than directly citing the original article text. Struggles were observed when participants were not sure which folder a certain article belonged to. For tags, more single words were used than multi-words. More tags were created by using the original article text than generalization. In the hybrid group, participants first decided which folder the current article should

belong to and then decided which tags should be assigned. Overall, hybrid participants tended to create a flatter folder structure than the folder group, but they tended to assign more tags than the tagging participants. Since the hybrid group participants used both folders and tags in the file organization session, they experienced the difficulties with creating folders and creating tags.

The folder structure and the tagging structure were not totally contrary to each other. Similar behavior across different groups was observed. In the experiment, we noticed that some tags created by participants in the tagging group were very similar to the folder names created in the folder group. These tag names were high-level generalizations to the content of the article and they were reassigned to additional articles over and over again.

5.2 Re-finding Files with Folders and Tags

How did participants re-find files with different given clues? What kind of cognitive process did they go through during the file re-finding session? This section will focus on these questions.

5.2.1 Previous Models on File Re-finding

5.2.1.1 Akin et al. (1987)'s Model on Information Re-finding in a Directory Space

Akin et al. (1987) conducted an experiment on users' search behavior in a directory environment under the UNIX operating systems. Compared with modern hierarchical folder systems, where users navigate through folders and subfolders by clicking on the icon of a folder using a mouse, a UNIX directory requires users to type in UNIX commands to traverse the structure. Akin et al. proposed a four step model for the re-finding process in the UNIX directory space: display, recognize, strategy and action. Users first used commands to display information on the screen and then recognized information displayed on the screen based on memory. They then took an action based on what they recognized on the screen, either moving on to new directories or narrowing down the search.

5.2.1.2 Pak et al.'s Re-finding Model

Besides an organization model, Pak et al. (2007) proposed a re-finding model for the folder structure and the tagging structure. In both the models for the folder and the tagging structure, users made selections of categories

or tags based on the concepts activated within users depending on the content they were wishing to find. The difference between the folder structure and the tagging structure was that there is an additional judgment step on judging subcategories that when interacting with folders.

5.2.1.3 A Recall and Recognition Process

Akin et al. (1987) found that users tended to try different directories or folders based on their recall of the files and then recognize these files in the UNIX system. Barreau (1995) found that users usually use the information that they can recall about a file to roughly decide where they should go to get access to the file. Then they browse through a list of files and find the desired file by recognition. Barreau called this behavior "location-based search." Although Fertig et al. (1996) argued that location-based search is no different from a user controlled keyword search, later studies suggested that location-based search and user controlled keyword search were different. Alvarado et al. (2003) found that people prefer to "use contextual information as a guide in navigating locally in small steps (pp.4.)" rather than directly using the keyword search. Boardman & Sasse (2004) reported similar findings to Alvarado et al. (2003). They found that participants preferred browsing over searching. Two related types of browsing were identified from the study, first being location-based browsing of folders and desktop icons, second being sorting or scanning of items. The re-finding behavior consisted of a recognition process following a recall process. Recall is the process by which participants come up with memory cues about the target articles. At this stage, users try to search their internal memory to find "knowledge in the head" about the desired files. Recognition is the process by which participants interact with the organizational structure and decide if the retrieved article is the to-be-found article. At this stage, users try to utilize "knowledge in the world" provided by the organizational structure and find the target files.

5.2.2 Re-finding Strategies Reported by Participants

After the file re-finding session, participants answered questions regarding the re-finding strategies in re-finding. Answers collected from the questionnaire are reported in this section.

5.2.2.1 The Folder Group

In the questionnaire, the folder group participants reported that they used similar approaches for the four re-finding tasks, although the challenging level increased from task 1 to task 4. Participants commented that the

strategies they used for the latter tasks were “the same process as in task 1”; “same as previous tasks”; “same thing [as previous tasks]”. Participants tried to remember which folder the files were saved by building connections between files and folders using the provided clues: article title, article summary, simple question and complex question. If they remembered where the particular article was saved, they directly opened the folder to recognize the target file. If they did not remember where the desired article was saved, they opened the most likely folder based on matching the provided clue with the folder structure they created. They checked folders and subfolders to narrow down the search. When the tasks became more challenging from task 1 to task 4, participants usually had to look into several folders before they re-found the target articles. Sometimes they had to read the file content to make sure that it was the one they were looking for. In task 4, the scenario when participants needed to re-find multiple files for a relevant topic, some of them combed through the entire folder structure to gather all the relevant articles.

Not only did participants try to recall what they remember about the articles using the given clues, they also tried to “go back in time,” meaning that participants were trying to think the same way that they did in the file organization session. For instance, one participant mentioned: “[In order to re-find the files,] I rationalized where I think I would have placed the file before [in the file organization session]”. When participants worked on the 60 articles in the file organization session, the entire article collection constructed a context that helped them to remember where they saved the files and provided reminding functions in the re-finding session.

5.2.2.2 The Tagging Group

For the tagging group, memory played an important role in the re-finding process as well. In an ideal situation, when participants remembered perfectly well what exact tags were assigned to a particular file, the re-finding process was more a matter of navigating to any of the tags. Since all of the tags were listed in a small area in the Zotero interface, navigating to a specific tag usually required participants scrolling up and down the tag list to pin point the particular tag.

In situations when participants did not remember exactly what tags they created for a target file, they browsed through the tag list and looked for tags that were semantically relevant to the clues provided. Browsing for the purpose of recognition was commonly used for re-finding under such circumstances. If they could not remember anything about the target file, they were more likely to browse through the tag list and click on those

tags that might be relevant to the clues and check if the articles associated with those tags were the ones they were looking for. In such situations, browsing the tag list provided reminding functions in the re-finding session.

5.2.2.3 The Hybrid Group

The hybrid participants had options of using both folders and tags for re-finding. They started the re-finding process by either clicking on a tag first or a folder first. Our previous discussion showed that there was no apparent preference in the first choice on a folder or a tag. Participants made the first choice based on whether it was easier to build a connection using the given clue with a folder or with a tag. For instance, "I wanted to get tags from the article summaries, but mostly that took too long. So I looked into folders in this part." In the task when the given clues were article summaries, a direct connection was easily built up between summaries and folders because usually folders were high level generalization of files contents.

Besides using either folders or tags exclusively for re-finding, participants reported that they also used a combination strategy of using both folders and tags. For instance, one participant reported: "I mostly went to the folder first, and then I picked a tag to narrow the list." The strategy was successful when the folders were general enough to contain a list of articles that can be further differentiated by tags.

5.2.3 Analysis of the Re-finding Process from the Screen Capture Videos

Besides analyzing the participants' answers from questionnaires, the researcher watched the screen capture videos closely in order to understand the re-finding process. By analyzing the videos, multiple stages that participants went through in the re-finding process were uncovered.

5.2.3.1 The Folder Group

Before the folder participants started clicking on any folders, they read the provided clues, including article titles, article summaries, or relevant questions. After that, they browsed through the organization scheme. In the browsing stage, what they tried to do was to build up a connection between the given clue and the previously created organization scheme. In order to build up the connection, they recalled the organization process in the previous session.

After the browsing stage, participants usually already had a sense of whether they were sure or not sure about where the file was. If they were confident about the intuition, they would directly open the folder or the

subfolder and then recognize the article. This circumstance usually happened when the re-finding task was easy or when the participant had a good memory of the particular file. However, what happened more often was that they were not sure about where exactly the file was. They hesitated and not sure where to start. They would try two or three most promising folders that they can connect with the clues and brows the files listed in those folders. If no files were recognized, they started to click on random folders and browsed the files listed in those folders. With more challenging re-finding tasks, participants experienced more struggles and even had to try a comprehensive search, by clicking on every single folder to see the files listed in them. Participants experienced stages from hesitance to random click, and to comprehensive search if the task was difficult.

Task 4 was different from the previous three re-finding tasks in that there was more than one desired article and usually these articles were located in more than one folder. Participants usually had to check on several folders in order to re-find all the relevant articles. We found that participants went through the similar stages as the previous tasks. They first checked the most likely folders, then the ones they thought might be promising, then random clicks on remotely related folders, and in some cases they combed through the entire folder structure in a comprehensive search to make sure they did not miss any relevant articles.

5.2.3.2 The Tagging Group

Before participants started clicking on tags, they read the given clues trying to recall what tags might be associated with the target files. After the participants read the given clues, they browsed through the tag list and clicked on those they thought would lead to the target files. If they were confident and they knew exactly which tags were associated to the target file, they browsed through the tag list and tried to locate those tags. After they located and clicked on those tags, they would recognize the file from the list of files that were associated with the tags. If they were not so sure about what tags they were looking for, they tried a few synonyms and semantically related tags. If no results were found, they simply started to try random clicks on any possible tags. Participants did not check every single tag because the number of tags was usually large.

Two major strategies were used by the tagging participants: a strategy of combining tags and a strategy of tag chain. Combining tags to narrow down the re-finding scope was a strategy widely used by participants. After participants clicked on one particular tag, articles that were associated with this tag were listed. Meanwhile, all other tags that were associated with these articles were listed in the tag area. Participants then selected one or

more tags from this shortened tag list to combine them with the previously selected tag for narrowing down the recognition file list. For example, a participant clicked on the tag “travel” to re-find a travel related article in task 3. Ten articles that were assigned the tag “travel” showed up in the middle pane of the Zotero interface. Meanwhile, 20 other tags that were assigned to these 10 articles were listed in the tag area, such as “airplane” and “motorcycle”. The participant then selected “motorcycle” to combine it with the previously selected “travel” and narrowed down relevant file list. At this point, the file that was assigned to both “travel” and “motorcycle” was the only file listed in the middle pane, which was exactly what the participant was looking for. In figure 19, the target file is file 3, users can re-find it by combining tag b and tag c.

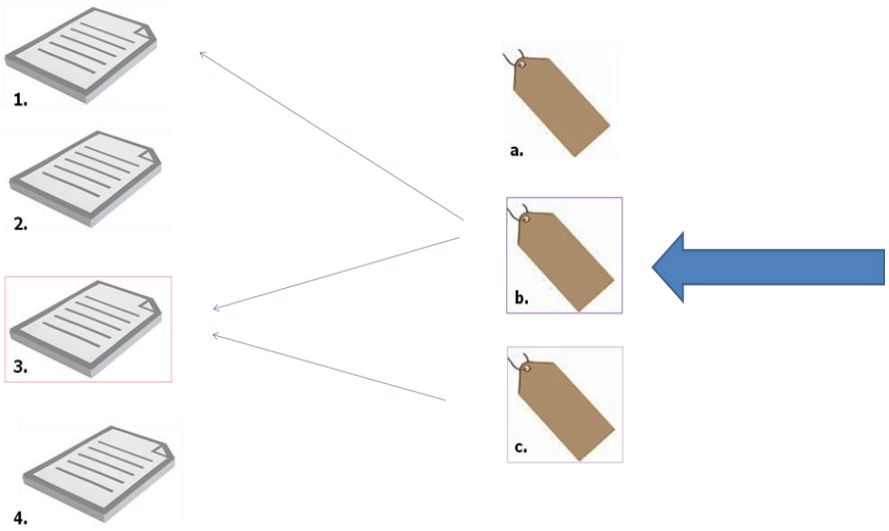


Figure 19. The Strategy of Combining Tags

Following a “tag chain” was another interesting strategy observed from the tagging group. By “tag chain” we mean that participants discovered the desired file by following a chain of tags that were connected by the files that the tags were assigned to. For instance, a participant selected “travel” to re-find a travel related article in task 3. Ten related articles and 20 related tags to these articles were listed, such as “airplane,” and “motorcycle.”

The participant then selected “motorcycle.” But instead of combining it with “travel”, he unselected “travel”. Now instead of seeing the files associated with both motorcycle and travel, the participant saw all the five articles associated with “motorcycle” listed in the file list. All the fifteen tags that were assigned to these five articles were listed in the tag list, for example, “accident,” “banned,” and so on. The participant then scanned through these 15 tags and selected “accident” to re-find an article that was about motorcycle accidents. In other words, the file was tagged with “motorcycle” and “accident” but not with “travel”. However, the first tag that the participant could think of was “travel”. He started with the tag “travel” anyway, but jumped to related tag of “motorcyle” when the more relevant tag showed up. In figure 20, the target file is file 3. Users can re-find the file by starting with tag a, and then following the tag chain and jumping to tag b.

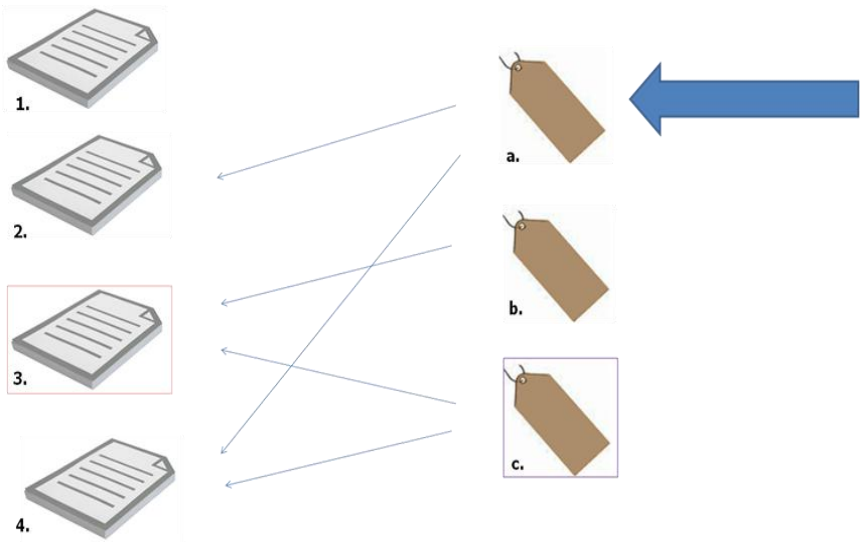


Figure 20. The Strategy of Following a Tag Chain

If the participant remembered what tags were associated with the article like in the example of “combination strategy”, the participant found the desired file by combining the tag of “travel” and the tag of “motorcycle”. If the participant did not remember any of the tags associated with the article, like in the example of “tag chain”, the participant found the desired file by starting with a tag of “travel” that was not associated with the file, then

jumping to the tag of "motorcycle" which was associated with the desired file. Files are semantically connected by sharing multiple tags. Participant can start with any tag that they thought of when they started the re-finding task. It does not matter if this is the "correct" tag that directly associated with the target file. The tag list provided a context that helped participant follow a memory trail and eventually found their way to the desired file.

In task 4, participants usually checked the most-likely tags first. The tag chain strategy was used extensively to bring out the relevant tags. Completely random clicking was also observed after the participants tried all the semantically related tags. The syntactic inconsistency of tag names (singular and plural forms, upper case and lower case, synonyms) made some articles completely invisible.

5.2.3.3 The Hybrid Group

Participants faced the choice of folder or tag in the beginning of the re-finding tasks and throughout the entire process. Overall, we did not see any apparent preference in the choice of folders or tags in the first click. For all re-finding tasks, slightly more than half of the re-finding cases started with a click on folders. Slightly less than half of the re-finding cases were started with a click on tags. We found that participants who clicked on a folder first were likely to also use tags for re-finding, and that participants who clicked on a tag first were likely to stick with tags for re-finding. In all re-finding tasks, about 40% cases where participants chose to use folder first also tried tags for the re-finding tasks. Only about 10% of the cases where participants chose to use tag first also tried folders for the re-finding tasks.

After they made the choice, the first step was similar to what we observed in the folder group and the tagging group. They either went directly to the most promising folders or the most promising tags. If the first step did not lead to the target file, the participants were faced with a selection again: sticking to the original choice, or turning to the other scheme for support, what we call the "combination of folder and tag" strategy. The combination strategy happened when participants first chose a general folder name that contained multiple articles, and then narrowed down the search by selecting detailed tags that were associated with the articles within the folder. In figure 21, the target file is file 3. Users can find it by combing the folder and tag a.

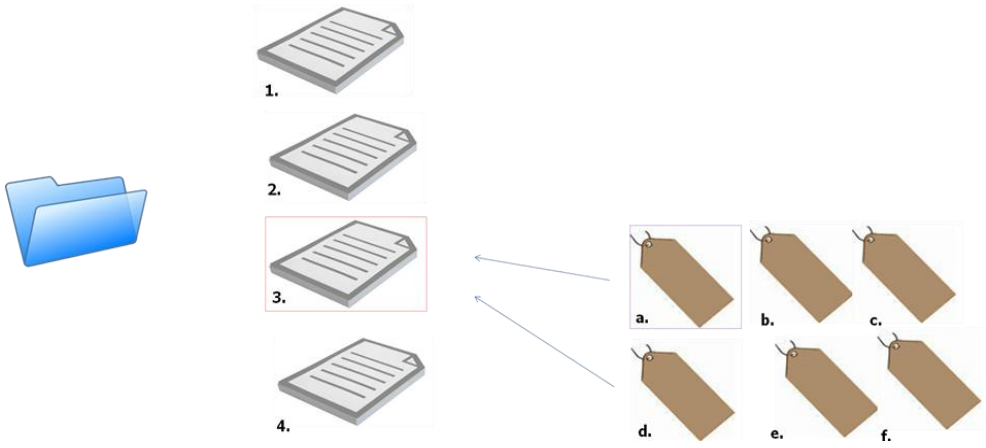


Figure 21. The "Combination of Folder and Tag" Strategy

Participants constantly had to decide which scheme to use when a mouse click was made and no correct target article was found. If participants stuck to their first choice, they continued re-finding files as if the other structure did not exist, until they reached a "switch point." A switch point took place when participants felt failure with the organization scheme (folder/tag) they were currently using, and they switched from one to the other (either from folder to tag or from tag to folder). The difference between switch point and combination strategy was that participants experienced failure with one organization scheme and decided to abandon this structure in the switch point, while in combination strategy participants were seeking complementation between the two structures.

There were three methods that the hybrid participants could use for re-finding: folder only, tag only, and use both folders and tags. We are interested in how often each of the three strategies was used in the re-findings tasks. Number of times each method was used was counted by observing the screen capture videos. No apparent preference was found. Using both folders and tags can be further differentiated into two situations: switching from one to the other and combining the two structures.

In task 4, we observed switching between folders and tags back and forth several times, which is viewed as a sign of struggle. This behavior was not observed in previous tasks. We also observed a conflict between the

folder scheme and the tag scheme in the re-finding process of task 4. A conflict between the folder scheme and the tag scheme happened when files in different folders share the same tags, as illustrated in figure 22. File 3 and file 5 reside in different folders but are associated with the same tag.

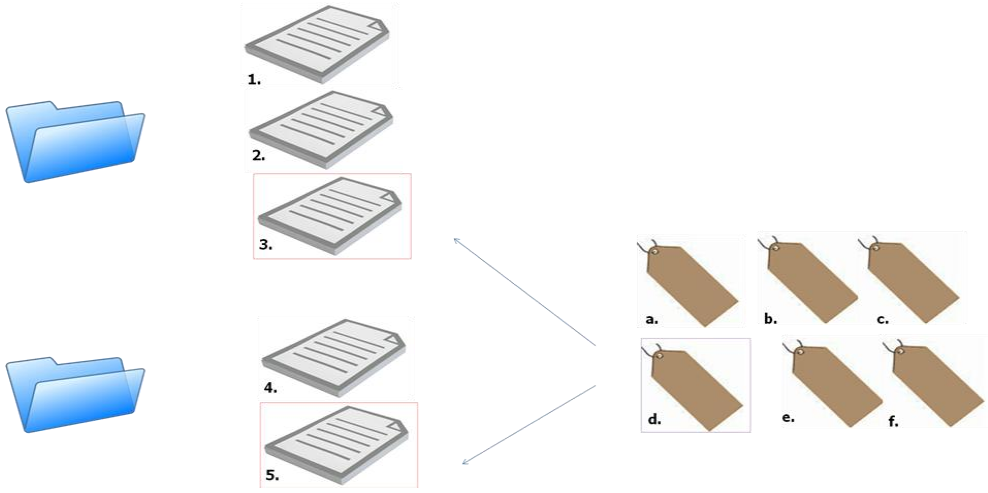


Figure 22. Conflict between Folders and Tags

5.3 Conclusion of Qualitative Observation and Discussion

This chapter qualitatively discussed the results of using folders and tags for desktop file management. The discussion was divided into two parts: a qualitative discussion on file organization with folders and tags, and a qualitative discussion on file re-finding with folders and tags. For file organization, we realized that organizing files on personal desktop is essentially a personal behavior. We saw differences in the organization behaviors both between the three the groups and within the groups. Nevertheless, similar patterns were observed in terms creating folders and tags. More folder names were created by generalization than directly citing the original article text. More tags were created by using the original articles text than high level generalization.

For file re-finding, in order to understand scenes behind the folder clicks and tag clicks, researchers watched the video recordings from the study and analyzed retrospective answers from participants. It revealed multiple re-finding strategies used in different groups. The tagging participants, in particular, used two interesting strategies

for file re-finding: a tag combination strategy and a tag chain strategy. These two strategies were used in different circumstances, either when they remember the tags associated with the desired file, or when they did not remember the exact tags associated with the desired file. In the hybrid group, we observed the conflict between the folder structure and tag structure that caused confusion in the re-finding process.

Chapter 6: Discussion and Conclusion

This chapter will discuss and conclude the dissertation. We will first answer the three research questions that we proposed in the beginning of the dissertation. We will then discuss the limitation of the research, the contribution of the research, and future work.

6.1 To Answering the Research Questions

6.1.1 To what extent does tagging support file management for file organization and file re-finding?

A simple answer to this question is yes, tagging does support file organization and file re-finding. The tagging structure required similar time in the file organization session, but required less number of mouse clicks in the re-finding tasks. We did not see any advantages of the tagging structure in other measurements such as file re-finding time and answer correctness.

71.4% of the participants in the tagging group were first time tag users. They did not have any previous experience with tags. In the tutorial session, they learned about what tags are, how tags could be added in the Zotero system and how tags could be used for file re-finding. The experiment showed that they had no difficulty using a tag-only structure for file organization and file re-finding. The experiment also showed that participants spent approximately equal time in organizing the files using folders and tags. Since generally more tags were created than the number of folders assigned, it confirmed our predication that creating single tags is cognitively easier than assigning a file to a folder.

We observed the necessity of revising the folder structure at the end of the file organization session, but not so much in the tagging structure. The folder participants complained that they would have done a better job in organizing if they had been given a chance to browse through the entire article collection in the beginning. An overview of the entire article collection would allow them to create a better folder structure that works for all the files, while the tagging participants did not have such complaints. Although some tagging participants revised the tags that they assigned to previous articles based on what they read later on, such revision did not require a high level overview of the entire article collection.

Compared with the folder group, tagging participants used a similar amount of time in completing the re-finding tasks, and achieved a similar level of answer correctness, and what interested us most was that they used less number of mouse clicks to get access to the desired files. As tagging participants seemed to be comfortable without folders, some of them suggested a "tagging folder" or a "tagging hierarchy", meaning that tags should be presented in a hierarchical way instead of a simply alphabetical manner. Results from the re-finding session suggested that the tagging structure could provide as much re-finding support, if not better, as the folder structure. As we discussed in the literature review section, deep hierarchical folders make files inside hierarchies invisible. Tagging structure provided more visibility by allowing users to get access to files with any assigned tags. Tagging structures act as a flat hierarchy that provides more access points.

Combining tags to narrow down the re-finding scope was a strategy used by participants. After participants clicked on one particular tag, all the articles that were associated with this tag were listed. Meanwhile, all other tags that were associated with these articles, rather than the entire tag list, were listed in the tag area. Participants then selected one or more tags from this shortened tag list to combine with the previously selected tag for narrowing down. Following a "tag chain" in selecting tags was another interesting strategy used by the tagging group. Tag chain was used when participants did not start with a tag that was directly associated with the target file. The ability to combine tags and follow a tag chain provided more flexibility for re-finding.

In summary, our study showed that using tagging for file organization and file re-finding achieved similar effectiveness and efficiency to using hierarchical folders. The only advantage we found from the study was that using tagging required less number of mouse clicks than folders to re-find files.

There could be several reasons for it. First, the display of tags made it hard for users to locate a specific tag even if they knew exactly what tag they were looking for. Zotero had a reserved space on the left lower corner of the interface for the tag list. All tags were listed alphabetically in the reserved space. In cases when there were more than 100 tags in the limited space, it was very time consuming to scroll up and down to find a specific tag. Second, uncontrolled vocabulary in creating tags resulted inconsistency in tag names. Participants used different words to describe the same tag. As Golub et al. (2009) found in their study of Entag, controlled vocabulary could be useful for both organization and re-finding, to help produce ideas of tags to use, to make it easier to find focus for tagging, as well as to ensure consistency and to increase the number of access points in retrieval.

6.1.2 To what extent does combining tagging and folders support file management for file organization and file re-finding?

A simple answer for this question is no, combining tagging and folders does not support file organization and file re-finding as well as we expected. Compared with the folder and the tagging group, the hybrid group spent longer time in the file organization session, but did not show any advantages in re-finding files in any of the measurements: file re-finding time, answer correctness and number of mouse clicks. Since the hybrid structure provided both folders and tags for organization, researchers required participants to use both of them for organization. This arrangement resulted that participants created more detailed organization schemes than either one used alone. Such organization schemes did not improve efficiency and effectiveness in the re-finding process. We believe the reason that caused such results in the hybrid group is the independent nature of both the folder structure and the tagging structure. The participants were given both of the two structures and they spent as much effort as they wanted in creating the organization schemes. Although our expectation of an ideal hybrid structure was a high-level folder and more detailed tags within folders, this was not observed in all participants' behavior. They ended up creating detailed folders and tags and sometimes conflicts between the two structures existed.

From our literature review, we learned that hierarchical folder structure and tagging structure each has its advantages and disadvantages in file organization and file re-finding. We expected that a hybrid system with both hierarchical folders and tags would partially solve the deficiencies of each structure when used alone. Users first used high-level hierarchy to categorize information into different topics. Users then used tags to help manage information if more detailed and complicated categorization is needed. However, our detailed analysis into the re-finding strategies used by hybrid group participants showed that only very few people combined using folders and tags in the way that we expected. Using folders exclusively, using tagging exclusively, and using both folders and tags, were each used in about one third of the re-finding tasks. The case of using both folders and tags could be further differentiated into two sub-cases: combining the use of folders and tags, and switching from one structure to the other. In the combining cases, participants used the hybrid structure the way as we expected: folders were used to roughly decide the topic and tags were used to further pin point the desired article. In the switching cases, participants felt failure with one organization scheme that was currently in use

and decided to switch to the other scheme. A "switch point" is when participants felt failure with the organization scheme (folder or tag) they were currently using, and switched from one to the other (folder to tag or tag to folder). We also observed conflicts between the folder scheme and the tag scheme in the re-finding process. A conflict between the folder scheme and the tag scheme occurs when files in different folders share the same tags. Such confusion caused more struggles in the re-finding process.

Another possible reason for the unsatisfactory results of the hybrid group is the Zotero interface we used in our experiment. Zotero simply put both folder and tag on the same interface, without providing any further integration function. Although there is possibility that the advantages of the folder structure and the tagging structure could be combined, it was not achieved in our experiment.

6.1.3 What are users' preference among the three structures, hierarchical folders, tagging, and the hybrid structure?

The questionnaires that we designed did not provide a good answer to this question. Each participant only used one structure for the organization and re-finding tasks. No matter which structure participants used, they found sufficient support for organization and re-finding, and experienced struggles and troubles in the process. Therefore, we did not see any significant difference in the comparison of post-task preference from participants, in both the file organization and the file re-finding session. A different experimental design that asks one participants to work on different organization structures might produce different results.

The hybrid participants were the only group who had access and experience with both the folder structure and the tagging structure. We expected that they would provide a better feedback on which structure provided more support for file organization and file re-finding. In analyzing the results from the questionnaire, we found that the hybrid participants believed it was easier to create tags than to create folders. We also found that the hybrid participants gave a lower helpfulness score to folders than the folder participants gave to folders. However, we did not see a preference of tags over folders in the hybrid group. Participants made the choice of using folder or tag based on which one was more "clickable" at the point. If the participant built a better connection between the given clue and one of the folders, then the folder structure was more "clickable" at that point. If the participant built a better connection between the given clue and one of the tags, then the tagging structure was more "clickable" at that point. After they made the first choice of either starting with the folders or with the tags,

they would switch back and forth based on the changing "clickability". If the folders they created for a certain file did not provide enough support for re-finding, they might switch to tags for help. Similarly, if the tags they created for a certain file did not provide enough re-finding support, they simply switched to folders. Since both the folder structure and the tagging structure helped them in re-finding, it was not surprising that we did not see any preference between the two in the questionnaires.

6.2 Contribution of the Research

This research aimed at exploring the possibility of incorporating tagging into the current hierarchical folder structure for file management, especially for file organization and file re-finding. The primary contribution of the study is a comparison of three file management structures in the context of the desktop file management: a hierarchical folder structure, a tagging structure, and a hybrid structure with both hierarchical folder and tagging functionalities. Advantages and disadvantages of each structure were revealed in the study.

The research also presented detailed analysis into users' file organization behaviors and file re-finding behaviors. When given different organization structures, users employed different strategies in organizing files and created different organization schemes for re-finding purpose. Multiple re-finding strategies used in the re-finding process were investigated. The research provided better understanding into how people organize and re-finding information. It shed light on how organization structure could affect the way people manage information.

This research has implications to the design of future document management tools. Hierarchical folders have been used for file management before the digital era and in the computing world. People have both positive and negative feedback about it. Emerging social technologies introduced new ways of organizing information on the web. They have impacts on the way people organize information and access information in desktop environments. This research provided empirical evidence that tagging structure could be used as efficiently as hierarchical folders for file organization and file re-finding. Although multiple tags allow users to re-find files from multiple paths, the study told us that this didn't resulted in higher correctness or shorter amount of time in re-finding. It seemed each structure has its own drawbacks in terms of helping people organize and re-find files. Although our attempt in trying to make use of both the advantage of folders and tags did not work out, we believe that a system that provides better support for using folders and tags together might help us solve this problem. The Zotero interface we used was not flexible enough to create and maintain both folders and tags.

6.3 Limitations of the Research

6.3.1 Contextual Tags

One fundamental characteristic of real life information is the context and environment in which the information was created, used, remembered and retrieved. The complexity of contextual information could not easily be captured in a laboratory experiment. In our experimental design, we put the 60 articles in the same environment with the same context. As we expected from the literature review, one advantage of using tags is that subtle information that might not be easily captured in a folder name could be put into a tag, such as “important,” and “has to read”. File management is a longitudinal process itself with complex environments changing over time. A file that was tagged as “important” yesterday could be no longer important for today. In our experiment, we specifically eliminated the consideration of contextual information in order to compare the three structures in terms of content organization. A study that puts folders and tags into real life setting might reveal more details on how folders and tags could be used for file management.

6.3.2 The Use of Zotero

Zotero is a free firefox extension that was developed for the purpose of collecting and managing citations. We customized the software into three versions for our study. Participants who were recruited for the study did not have any previous experience with the software. We therefore conducted a tutorial session before the file management session started. Although Zotero itself was not difficult to use, we observed troubles and problems when participants were interacting with it. Sometimes they clicked on the wrong button, and some other times they accidentally closed the software. Such incidents resulted in longer time in the organization and re-finding process, and added more struggles on top of the struggle participants had with the folders and tags. In the context of desktop file management, users are usually very familiar with the tools and software they install and use on their own computer. This is one of the limitations of using Zotero for our study.

The other limitation of using Zotero is the capability of the system in combining the use of folders and tags. As we discussed previously, the system simply put folder and tag in the same interface, and allows users to create both folder and tag for files. It did not provide a good way to help users to make the best of both structures. We observed that users created detailed folder structure and tagging structure in the hybrid group, either one of

which could be used independently for organization and re-finding. Also, the dual structure of detailed folders and tags did not help the participants re-find files easier, but rather caused confusion in the re-finding process.

6.3.3 More Experimental Design Considerations

The experiment was designed to collect both quantitative data and qualitative data. We collected the following quantitative data: time to complete file organization, time to complete file re-finding, number of mouse clicks, number of correct answers and incorrect answers. The qualitative data that we collected included the following items: participants' self-created organization schemes, participants' interaction with the interface recorded by MORAE recorder, and questionnaires answered by participants. The whole data set provided an in depth view of the study. However, we think that data collection could be enhanced if the following was taken into consideration. First, we believe that a think-aloud session in the file re-finding session could be more helpful than the retrospective questionnaire answers. Participants were usually very brief in writing in terms of what kind of strategies they used in the re-finding process. They sometimes did not remember the subtle differences in completing the four different re-finding tasks. It would be more useful to catch the differences while they were in the experiment, instead of waiting until they finished the entire session.

6.4 Future Work

Large amounts of data were collected from the dissertation project. More analysis can be done on the data, beyond the scope of my dissertation. One analysis that we plan to do is to map each groups' organization strategy using a co-folder and co-tagging analysis. Also, the dissertation only studied users' organization and re-finding with textual documents. To investigate the advantages of using tags for file management, file management on multiple formats of information is an area that we want to work on in the future. In the dissertation research, we explicitly disabled the use of a search engine for the re-finding tasks. We would like to explore the possibility of incorporating search engine with tagging systems. A longitudinal study into the users' own personal files and how tags can be used to enhance personal information organization and re-finding might also be interesting. Tagging systems that enables automatic tag generation is another area that we would like to work in the future.

List of References

Abrams, D., Baecker, R., and 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.

Alvarado, C., Teevan, J., Ackerman, M. S., and Karger, D. (2003). Surviving the information explosion: how people find their electronic information. Technical report AIM-2003-006. Massachusetts Institute of Technology, April 2003.

Akin, O., Baykan, C., and Rao, D. (1987). Structure of a directory space: A case study with a UNIX operating system. International Journal of Man-Machine Studies 26: 361- 382.

Agosto, D., Ma, S., Bui, Y. and Westbrook, L. (2007). A Mixed-Methods Study of the Use of Qualitative Research Methods in Recent Library & Information Science (LIS) Research. Library & Information Science Research (In Press).

Bälter, O. (2000). Keystroke level analysis of email message organization. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 105-112.

Barreau, D. K. (1995). Context as a factor in personal information management systems. J. Am. Soc. Inf. Sci. 46, 5 (Jun. 1995), 327-339.

Barreau D. (2006). Personal Information Management in Context. A SIGIR 2006 Workshop.

Bauer, D. (2002). Personal information geographies. In CHI '02 Extended Abstracts on Human Factors in Computing Systems (Minneapolis, Minnesota, USA, April 20 - 25, 2002).

Barreau, D. K. and Nardi, B.A. (1995). Finding and reminding: file organization from the desktop. SIGCHI Bulletin 27(3)

Bauer, D., Fastrez, P., and Hollan, J. (2005). Spatial Tools for Managing Personal Information Collections. In Proceedings of the Proceedings of the 38th Annual Hawaii international Conference on System Sciences.

Bellotti, V., Ducheneaut, N., Howard, M.A, and Smith, I.E. (2003). Taking email to task: The design and evaluation of a task management centered email tool. Conference on Human Factors in Computing Systems (CHI 2003).

Bellotti, V. and Smith, I. (2000). Informing the design of an information management system with iterative fieldwork. Conference on Designing interactive systems (DIS 2000).

Berghel, H. (1997). "Email : the good, the bad and the ugly." Communications of the ACM 40(4): 11-15.

Bergman, O., Beyth-Marom, R., Nachmias, R.(2006), The project fragmentation problem in personal information management, CHI 2006.

Bergman, O., Beyth-Marom, R. & Rafi, N. (2003). The user-subjective approach to personal information management systems. Journal of the American Society for Information Science and Technology 54(9): 872-878.

Blanc-Brude, T. & Scapin, D. L. (2007). What do people recall about their documents?: implications for desktop search tools. In Proceedings of the 12th international Conference on intelligent User interfaces.

Bloehdorn, S., Gorlitz, O., Schenk, S., & Volkel, M. (2006). TagFS – Tag Semantics for Hierarchical File Systems. In Proceedings of the 6th International Conference on Knowledge Management (I-KNOW 06).

Boardman, R., Spence, R. & Sasse, M.A. (2003). Too many hierarchies? : the daily struggle for control of the workspace. HCI International 2003.

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.

Boardman, R. (2004). Improving tool support for personal information management. PhD thesis. Dept. of Electronic and Electrical Engineering, Imperial College London.

Bondarenko O., & Janssen, R.(2005). Documents at hand: learning from paper to improve digital technologies, CHI 2005.

Bostrom, F. (2006). Personal Information retrieval.
http://www.cs.helsinki.fi/u/linden/teaching/irr06/drafts/fredrik_bostrom_pir.pdf

Capra, R., Pinney, M., & Perez-Quinones, M. A. (2005). Refinding is Not Finding Again. Technical Report TR-05-10, Computer Science, Virginia Tech.

Carroll, J. M. (1982). Creative names for personal files in an interactive computing environment. International Journal of Man-machine Studies, volume 16, number 4, pp. 405-438.

Civan A, Jones W, Klasnja P, & Bruce H (2008). Better to Organize Personal Information by Folders Or by Tags?: The Devil Is in the Details. Proceedings of the American Society for Information Science & Technology (ASIS&T) Annual Meeting, October, 2008, Columbus, OH.

Cockburn, A. & McKenzie, B. (2002). Evaluating the effectiveness of spatial memory in 2D and 3D physical and virtual environments. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems: Changing Our World, Changing Ourselves.

Cutrell, E. (2006). Search User Interface for PIM. Personal Information Management – A SIGIR 2006 workshop: 32-35.

Cutrell, E., Dumais, S. T., & Teevan, J. (2006). Searching to eliminate personal information management. Commun. ACM 49, 1 (Jan. 2006), 58-64.

Cutrell, E., Robbins, D., Dumais, S., & Sarin, R. (2006). Fast, flexible filtering with phlat. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (Montréal, Québec, Canada, April 22 - 27, 2006).

Ducheneaut, N. and Bellotti, V. (2001). E-mail as habitat: an exploration of embedded personal information management. Interactions 8, 5 (Sep. 2001), 30-38.

Dumais, S., Cutrell, E., Cadiz, JJ, Jancke, G., Sarin, R., & Robiins, D.C. (2003). Stuff I've seen: a system for personal information retrieval and re-use. 26th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 2003).

Edmonds, K.A., Blustein, J., & Turnbull, D. (2004). A Personal Information and Knowledge Infrastructure Integrator. Journal of Digital Information. (May. 2004), Article no. 243.

Engadget. www.engadget.com

Eric T. F. & Fertig, S.J. (1995). Lifestreams: Organizing your electronic life. In AAAI Fall Symposium: AI Applications in Knowledge Navigation and Retrieval, November 1995. Cambridge, MA.

Feinberg, J. & Millen, D. R. (2005). IBM Watson Research Center. Project Dogear. http://domino.watson.ibm.com/bridge/research.nsf/242252765710c19485256979004d289c/1c181ee5fbcf59fb852570fc0052ad75, (accessed on May 2,2008).

Fertig, S., Freeman, E., & Gelernter, D. (1996). Lifestreams: an alternative to the desktop metaphor. In Conference Companion on Human Factors in Computing Systems: Common Ground (Vancouver, British Columbia, Canada, April 13 - 18, 1996).

Freeman, E. & Gelernter, D. (1996). Lifestreams: a storage model for personal data. SIGMOD Rec. 25, 1 (Mar. 1996), 80-86.

Fertig, S., Freeman, E., & Gelernter, D. (1996). "Finding and reminding" reconsidered. SIGCHI Bull. 28, 1 (Jan. 1996), 66-69.

Giampaolo, D. (1998). Practical File System Design with the be File System. 1st. Morgan Kaufmann Publishers Inc.

Golub, K; Jones, C; Lykke Nielsen, M; Matthews, B; Moon, J; Puzoń, B; Tudhope, D (2009) EnTag: Enhancing Social Tagging for Discovery. *Joint Conference on Digital Libraries (JCDL), Austin, TX, June 15-19.*

Gonçalves, D. & Jorge, J. A. (2004). Describing documents: what can users tell us?. In Proceedings of the 9th international Conference on intelligent User interfaces (Funchal, Madeira, Portugal, January 13 - 16, 2004). IUI '04. ACM Press, New York, NY, 247-249.

Henderson, S. (2004). How do people organize their desktops?. In CHI '04 Extended Abstracts on Human Factors in Computing Systems (Vienna, Austria, April 24 - 29, 2004). CHI '04. ACM Press, New York, NY, 1047-1048.

Henderson, S. (2004). Personal Digital Document Management. APCHI 2004, LNCS 3101, pp. 651-655, 2004. http://staff.business.auckland.ac.nz/staffpages/shenderson/docs/APCHI2004DoctoralConsortium.pdf

Henderson, S. (2005). Genre, Task, Topic and time: facets of personal digital document management. CHINZ 05, July 6-8, 2005 Auckland, NZ.

Jones, W., Dumais, S. & Harry, B. (2002). Once found, what then? : a study of "keeping" behaviors in the personal use of web information. 65th Annual Meeting of the American Society for Information Science and Technology (ASIST 2002), Philadelphia, PA, American Society for Information Science & Technology.

Jones, W. (2004). Finders, keepers? The present and future perfect in support of personal information management. http://www.firstmonday.org/issues/issue9_3/jones/index.html

Jones, W. P. and Dumais, S. T. (1986). The spatial metaphor for user interfaces: experimental tests of reference by location versus name. ACM Trans. Inf. Syst. 4, 1 (Jan. 1986), 42-63.

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 (Portland, OR, USA, April 02 - 07, 2005). CHI '05. ACM Press, New York, NY, 1505-1508.

Kaye, J.J., Vertesi, J., Avery, S., Dafoe, A., David, S., Onaga, L., Rosero, I. & Rinch, T. (2006). To have and to hold: exploring the personal archive. CHI 2006.

Khoo, C., Luyt, B., Ee, C., Osman, J., Lim, H., & Yong, S. (2007). How users organize electronic files on their workstations in the office environment: a preliminary study of personal information organization behavior. Information Research, Vol. 11 No. 2, January 2007.

Kipp, M. E. I. & Campbell, D. G. (2006). Patterns and Inconsistencies in Collaborative Tagging Systems : An Examination of Tagging Practices. In Proceedings Annual General Meeting of the American Society for Information Science and Technology, Austin, Texas (US).

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 (Cambridge, Massachusetts, United States, June 25 - 28, 1989).

Kwasnik, B.H. (1991). The importance of factors that are not document attributes in the organization of personal documents. Journal of documentation, Vol. 47 No. 4, December 1991, pp. 389-398

Lansdale, M. (1988). "The psychology of personal information management." Applied Ergonomics 19(1): 55-66.

Lansdale, M. W. (1991). Remembering about documents : memory for appearance, format, and location. Ergonomics 34(8): 1161-1178.

Malone, T. W. (1983). How do people organize their desks?: Implications for the design of office information systems. ACM Trans. Inf. Syst. 1, 1 (Jan. 1983), 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 (Monterey, California, United States, May 03 - 07, 1992).

Marshall, C.C. & Bly, S. (2005). Saving and using encountered information: implications for electronic periodicals. CHI 2005

Marsden, G., Cairns, D.E. (2003). Improving the usability of the hierarchical file system. 2003 SAICSIT

Morae. www.morae.com

Mozilla. www.mozilla.com

Nardi, B. and Barreau, D. K. (1997). ""Finding and reminding" revisited : appropriate metaphors for file organization at the desktop." SIGCHI Bulletin 29(1).

Norman, D. (2002), The Design of Everyday Things, Basic Books, New York, NY, .

Jones, W., Bruce, H., & Dumais, S. (2001). Keeping found things found on the web. In Proceedings of the Tenth international Conference on information and Knowledge Management (Atlanta, Georgia, USA, October 05 - 10, 2001).

Quan, D., K. Bakshi, K., Huynh, D. & Karger, D.R. (2003). User interfaces for supporting multiple categorization. Interact 2003.

Robertson, G., Czerwinski, M. & Larson, K. (1998). Data Mountain : using spatial memory for document management. 11th Annual ACM Symsposium on User Interface Software and Technology (UIST 1998), San Francisco, CA, ACM SIGCHIACM SIGGRAPH.

Rekimoto, J. (1999). TimeScape: A Time Machine for the Desktop Environment. Human Factors in Computing Systems (SIGCHI).1999. Pittsburgh, PA: 180-181.

Sinha, R. (2005). A Cognitive Analysis of Tagging. http://www.rashmisinha.com/archives/05_09/tagging-cognitive.html

Spurgin, K.M. (2006). The sense-making approach and the study of personal information management. Personal Information Management – A SIGIR 2006 workshop: 102- 104.

Tagg. www.taggtool.com

Teevan, J. (2004). How people re-find information when the web changes. http://haystack.lcs.mit.edu/papers/teevan.aim04.pdf

Voida, S., & Greenberg, S. (2009). WikiFolders: augmenting the display of folders to better convey the meaning of files. CHI 2009: 1679-1682.

Whittaker, S. & Sidner, C. (1996). Email overload: exploring personal information management of email. Conference on Human Factors in Computing Systems (CHI 1996), Vancouver, B.C., ACM SIGCHI.

Whittaker, S. & Hirschberg, J. (2001). The character, value and management of personal paper archives. ACM Transactions on Computer-Human Interaction 8(2): 150-170.

Zotero. www.zotero.com

Appendix A: Tutorial

A-1: Tutorial for the Hierarchical folder Group

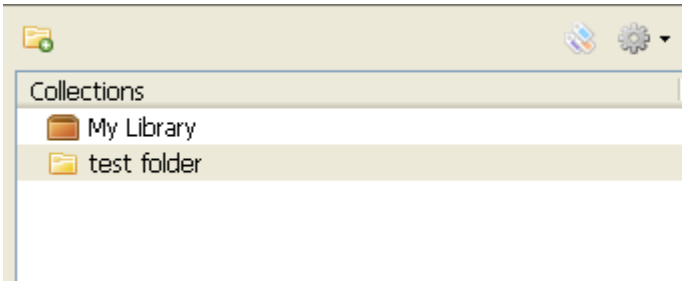
Tutorial

The window is divided into two sections: The upper level of the window shows you the content of files in the format of web pages: a page from Drexel ischool website as an example. The lower level of the window is Zotero's working space. Zotero allows you organize files with hierarchical folders. Zotero's working space is divided into 3 panes: left, middle and right. The left pane shows the organization structure. The middle pane shows the files that are saved. The right pane can be ignored.

How to create a folder?

1. Click on the folder icon on the left top of Zotero interface.

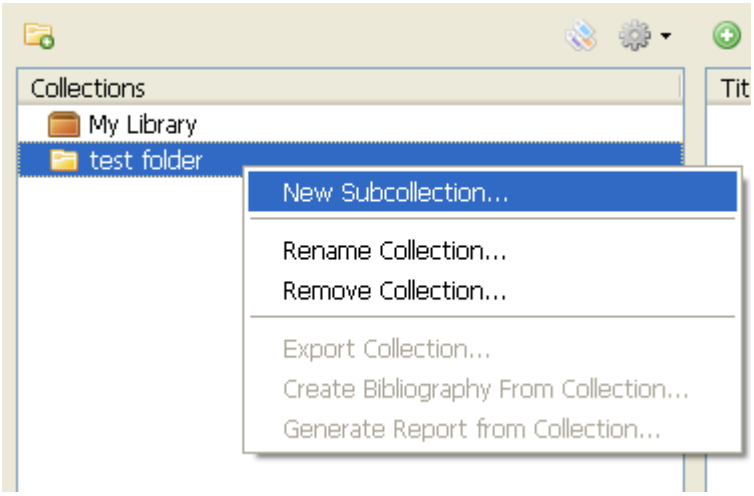
2. A box pops up in the middle of the screen. Type in the folder name you want to give to this folder: test folder. A "test folder" is created.



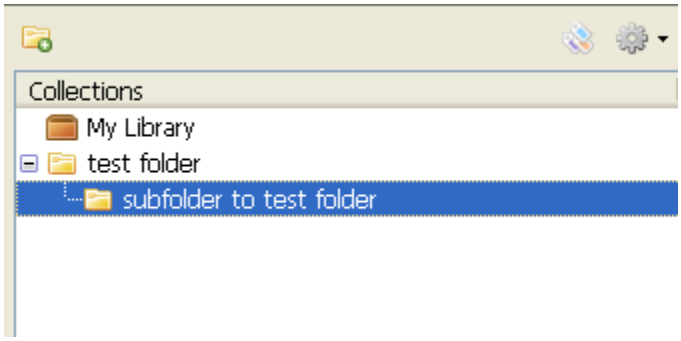
3. All the folders created this will be directly under "My library". Folders will be listed alphabetically.

How to create a new subfolder?

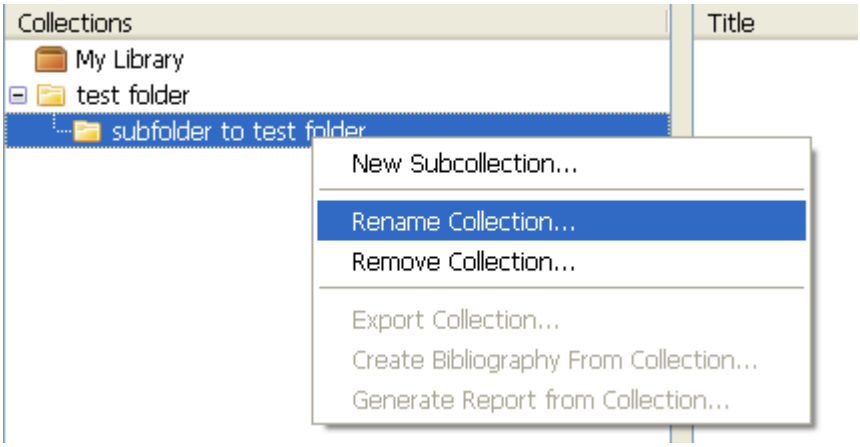
1. Right click the folder that the new subfolder belongs to. Select the option of "New Subcollection".



2. A box pops up in the middle of the screen. Type in the subfolder name “subfolder to test folder”. A new subfolder is created under test folder.



4. To rename or remove folders, right click the target folder to see the option.



How to save a file to a folder?

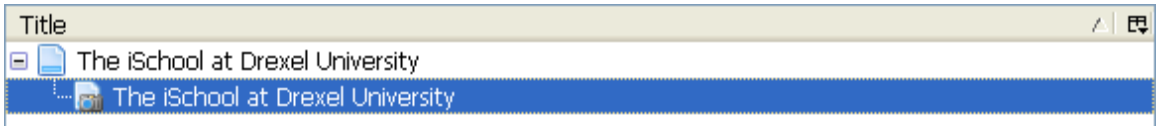
1. Select the target folder. This folder will be highlighted.

2. Click on the save file icon in the middle of the screen as shown below.

3. The current file is saved under the selected folder. File name is by default the name of the web page.

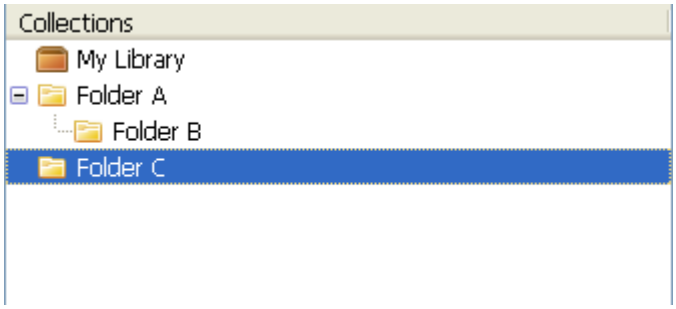
How to open a previously saved file?

1. Select the folder that the file is saved.
2. Select the file you want to open.
3. Click on the "+" sign in front of the file title.
4. Double click on the item shown below the original title. The saved file will be shown on the upper level of the window.

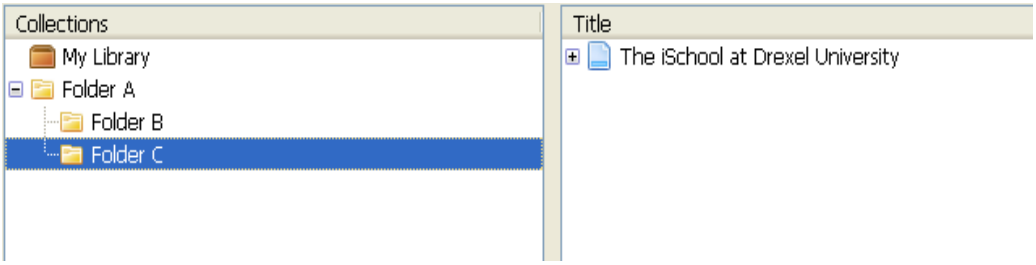


How to change the folder structure?

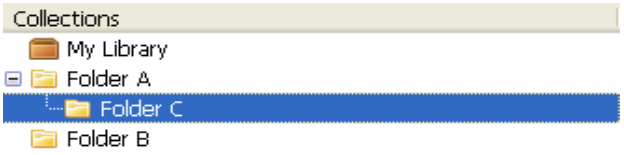
1. Delete the previous folder structure and files. Create a folder structure like this.



2. Save the current web page to Folder C.
3. Select Folder C, drag it to Folder A, and release it when Folder A is highlighted.
4. Folder C is now listed under Folder A. The File saved in Folder C is also moved together with the folder.



5. Folder B is originally listed under Folder A. To make Folder B no longer a subfolder to Folder A, select Folder B and drag it to "My Library" until "My Library" is highlighted.
6. Folder B is now listed separately from Folder A.

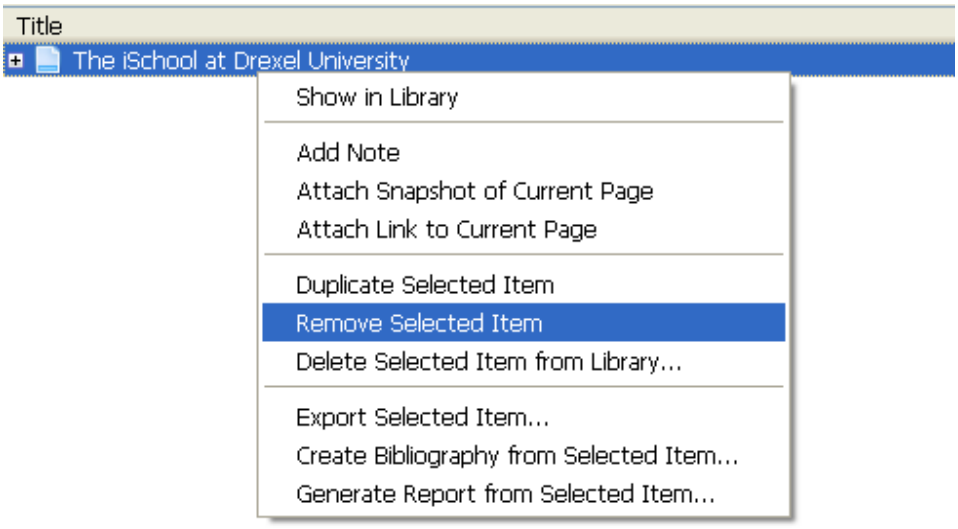


How to change a file's location?

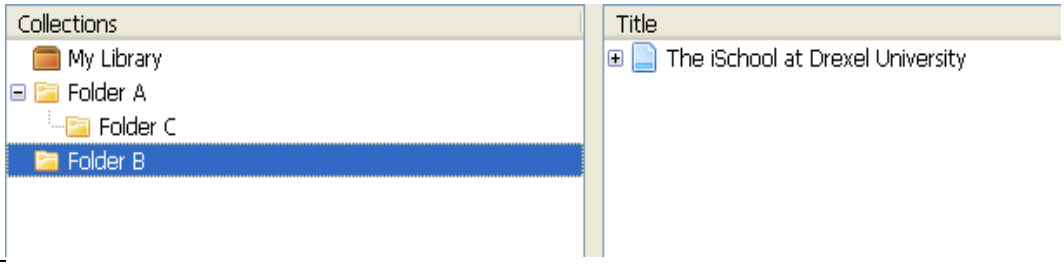
At this point, file "The iSchool at Drexel University" is saved in Folder C. Please follow the steps to move it from Folder C to Folder B.

1. Select Folder C, file "The iSchool at Drexel University" is shown in the middle pane.
2. Select the file, drag it to Folder B, and release it when Folder B is highlighted.

3. The file is now in both Folder C and Folder B.
4. Select Folder C, and remove the file from Folder C. Please pay attention that you can't select "Delete Selected Item from Library…" because it will delete the file permanently from all the folders.



5. The file is now only shown under Folder B.



A-2: Tutorial for the Tagging Group

Tutorial

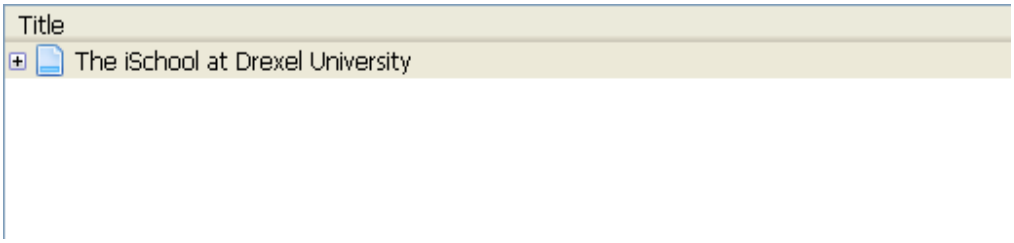
The window is divided into two sections: The upper level of the window shows you the content of files in the format of web pages: a page from Drexel ischool website as an example. The lower level of the window is Zotero's working space. Zotero allows you organize files with tags. Zotero's working space is divided into 3 panes: left, middle and right. The left pane shows the organization structure. The middle pane shows the files that are saved. The right pane shows the details of selected file.

There is a separate sheet on your desk talking about what tag is. Please read that page before you go further here.

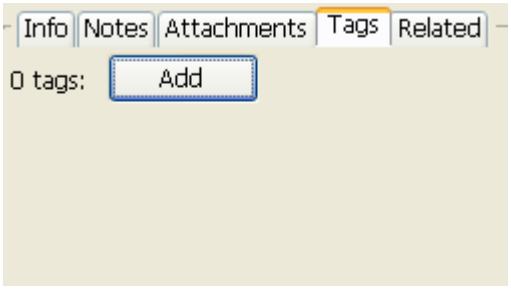
How to save a file and assign tags to it?

1. Click on the save file icon in the middle of the screen as shown below.

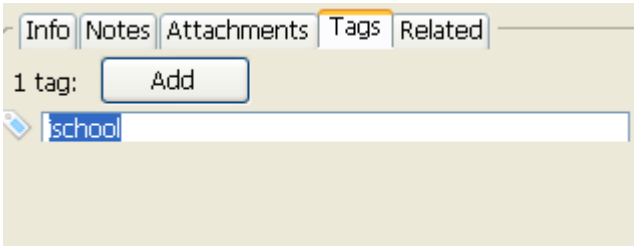
2. File is saved as shown in the middle of the screen. File name is by default the name of the web page. All files will be saved under “My Library”.



3. On the RIGHT pane, please click on the “tag” button.

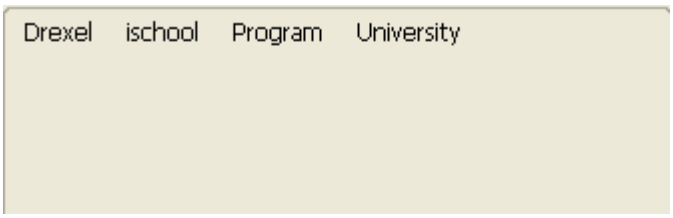


4. Click on “Add” to add tags to the file.



5. Assign more than one tag by clicking on “Add” again after you are done with the first tag, for example, Drexel, University, Program.

6. Tags to all saved files will also be shown on the LEFT pane, alphabetically.

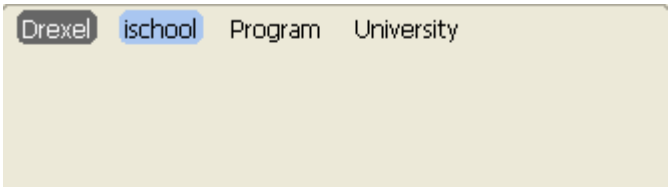


How to find your files using tags?

1. Click on the “ischool” tag on the LEFT pane, tags that are selected will be shown in blue. Files that are tagged with “ischool” will be shown in the MIDDLE pane, in this case, the ischool page we saved previously. All tags that are associated with ischool page will also be listed on the LEFT pane, in this case, Drexel, ischool, program, University.

2. Files can be found be by selecting multiple tags. To select more than one tag, click on other tags that you want to select. Files that are tagged with all selected tags will be shown in the MIDDLE pane, in this case, still the ischool page we saved previously.

3. To deselect one or multiple tags, click on the blue tags you no longer want to select.

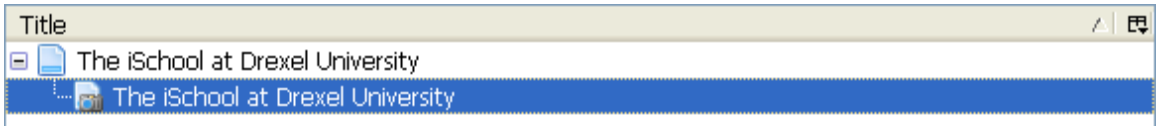


4. To deselect all the tags, click on the “Deselect” button at the LEFT bottom of LEFT pane, as shown below. None of the tags will be shown in blue after you deselect all.

How to open a previously saved file?

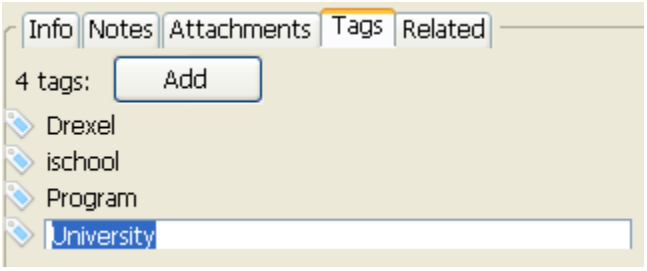
1. Select the folder that the file is saved.
2. Select the file you want to open.
3. Click on the “+” sign in front of the file title.

4. Double click on the item shown below the original title. The saved file will be shown on the upper level of the window.

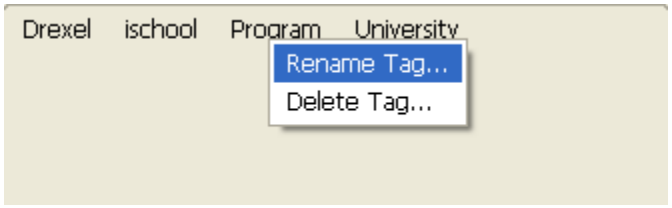


How to rename a tag?

1. To rename a tag for a single file, click on the tag on the RIGHT pane and start typing new tag name. Tag name will only be changed for this particular file.



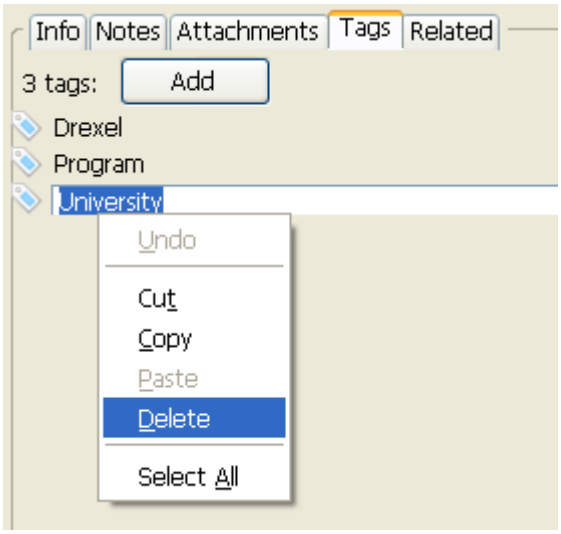
2. To rename a tag for all files that are associated with this tag, find the tag on the LEFT pane. Right click and select “Rename Tag”.



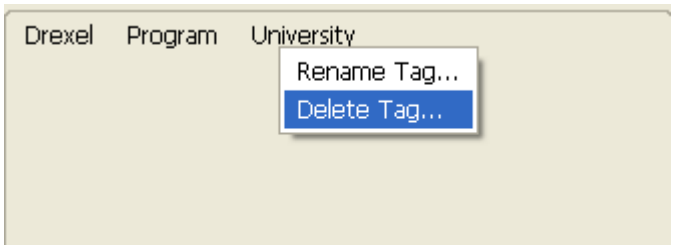
3. Type in the new tag name in the box and then click on OK. The tag will be renamed for all associated files.

How to delete a tag?

1. To delete a tag for a single file, find the target tag on the RIGHT pane. Right click and select “Delete”.



2. To delete a tag for all files that are associated with this tag, find the target tag on the LEFT pane.
3. Right click and select "Delete Tag". Then select Ok and the tag will be removed from all associated files.



A-3: Tutorial for the Hybrid Group

Tutorial

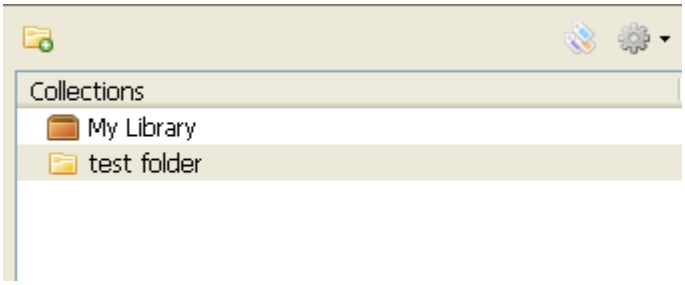
The window is divided into two sections: The upper level of the window shows you the content of files in the format of web pages: a page from Drexel ischool website as an example. The lower level of the window is Zotero's working space. Zotero allows you organize files with both hierarchical folders and tags. Zotero's working space is divided into 3 panes: left, middle and right. The left pane shows the organization structure. The middle pane shows the files that are saved. The right pane shows details of selected file.

There is a separate sheet on your desk talking about what tag is. Please read that page before you go further here.

How to create a folder?

1. Click on the folder icon on the left top of Zotero interface.

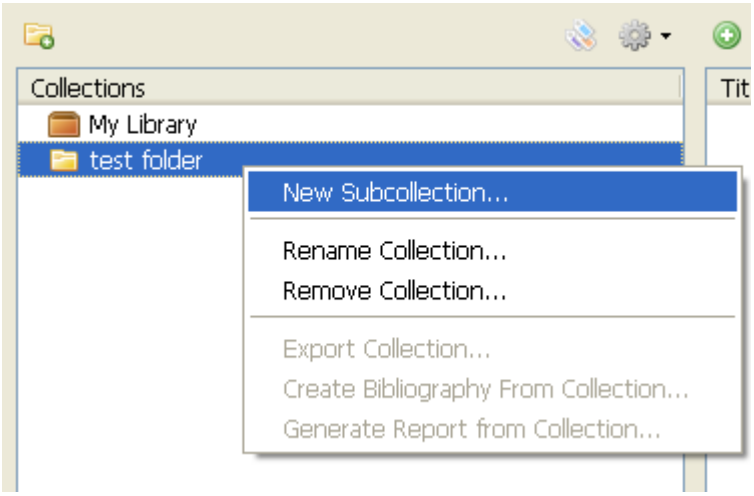
2. A box pops up in the middle of the screen. Type in the folder name you want to give to this folder: test folder.



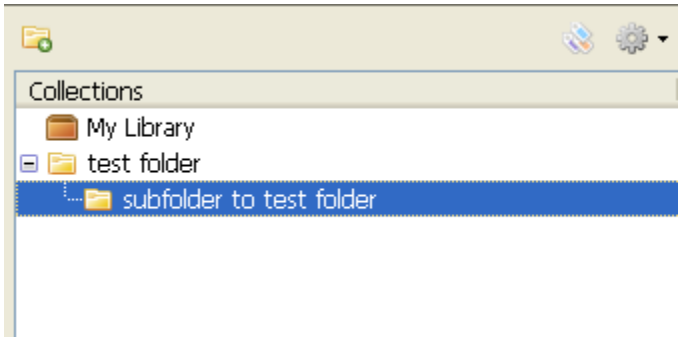
3. All the folders created this way will be directly under “My library”. Folders will be listed alphabetically.

How to create a new subfolder?

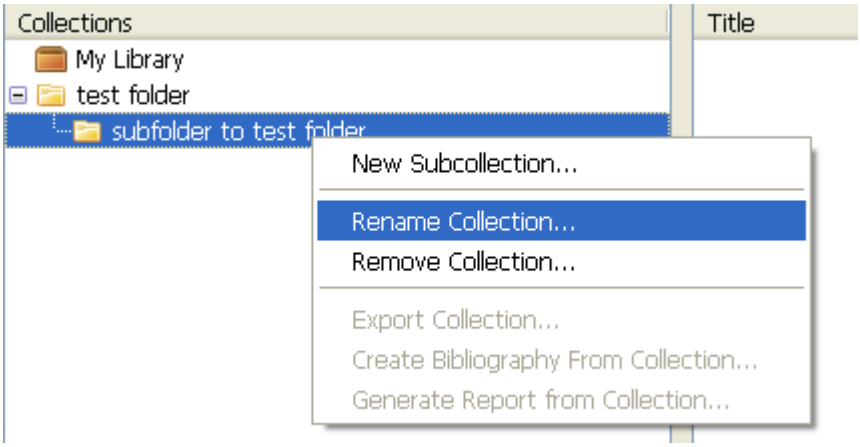
1. Right click the folder that the new subfolder belongs to. You will see the option of “New Subcollection”.



2. A box pops up in the middle of the screen. Type in the subfolder name “subfolder to test folder”. A new subfolder is created under test folder.



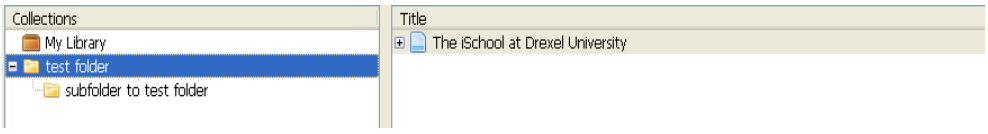
3. To rename or remove folders, right click the target folder to see the option.



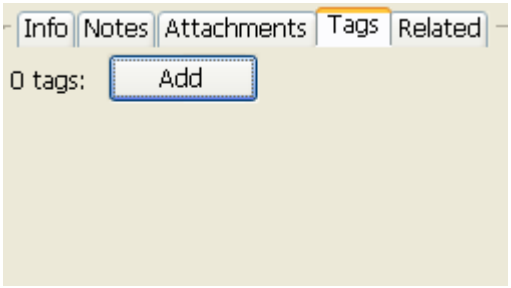
How to save a file to a folder and assign tags to it?

1. Select the target folder. This folder will be highlighted.
2. Click on the save file icon in the middle of the screen as shown below.

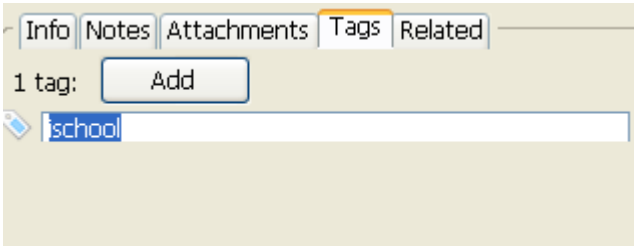
3. The current file is saved under the selected folder. File name is by default the name of the web page.



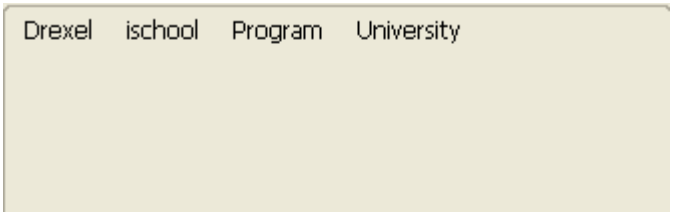
4. On the RIGHT pane, please click on the “Tag” button.



5. Click on “Add” to add tags to the file.

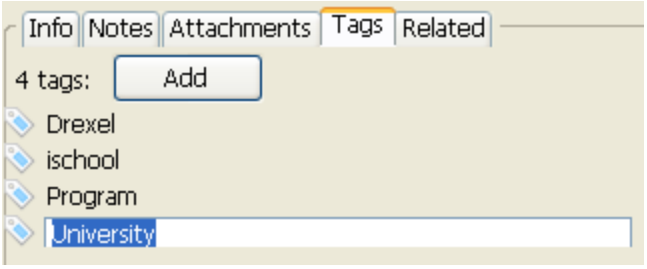


6. Assign more than one tag by clicking on “Add” again after you are done with the first tag, for example, Drexel, University, Program.
7. Tags to all saved files will be shown on the LEFT pane, alphabetically.

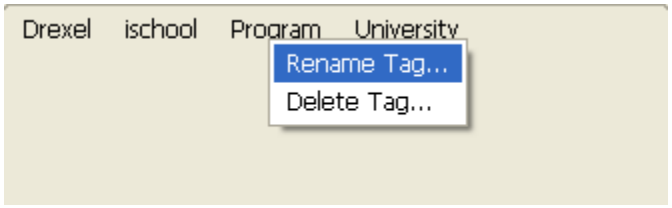


How to rename a tag?

1. To rename a tag for a single file, click on the tag on the RIGHT pane and start typing new tag name. Tag name will only be changed for this particular file.



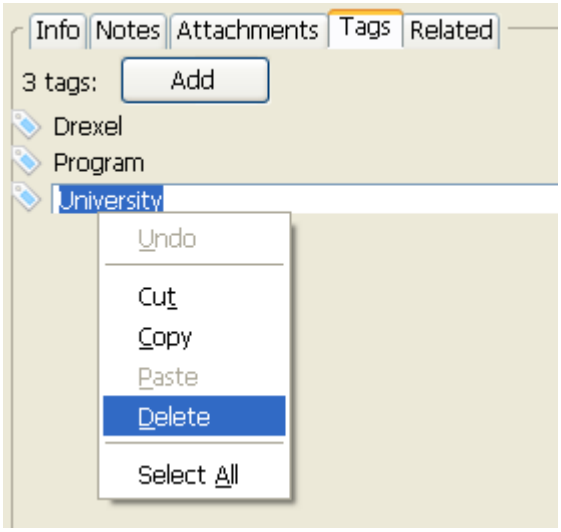
2. To rename a tag for all files that are associated with this tag, find the tag on the LEFT pane. Right click and select “Rename Tag”.



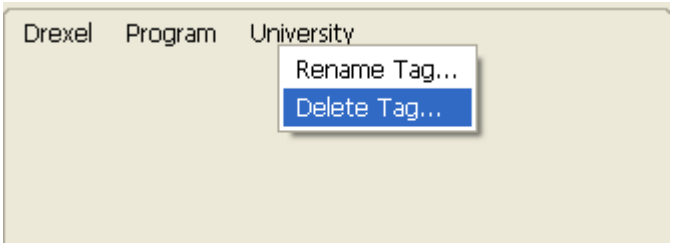
3. Type in the new tag name in the box and then click on OK. The tag will be renamed for all associated files.

How to delete a tag?

1. To delete a tag for a single file, find the target tag on the RIGHT pane. Right click and select “Delete”.

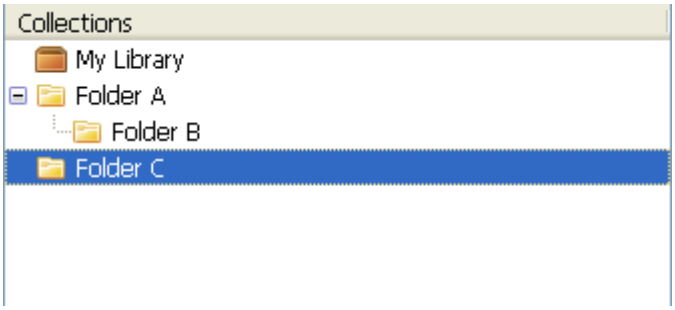


2. To delete a tag for all files that are associated with this tag, find the target tag on the LEFT pane.
3. Right click and select “Delete Tag”. Then select Ok and the tag will be removed from all associated files.

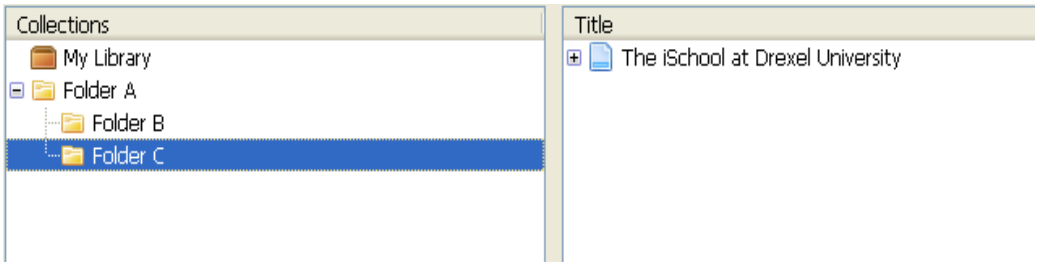


How to change the folder structure?

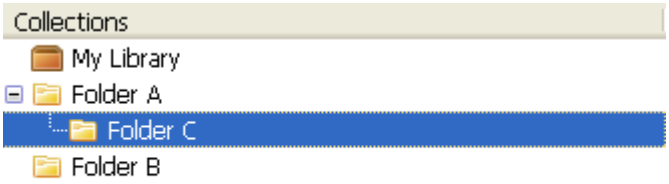
1. Delete the previous folder structure and files. Create a folder structure like this.



2. Save the current web page to Folder C and assign 4 tags to it: iSchool, Drexel, University, Program.
3. Select Folder C, drag it to Folder A, and release it when Folder A is highlighted.
4. Folder C is now listed under Folder A. The file saved in Folder C is also moved together with the folder.



5. Folder B is originally listed under Folder A. To make Folder B no longer a subfolder to Folder A, select and drag Folder B to "My Library" until "My Library" is highlighted.
6. Folder B is now listed separately from Folder A.

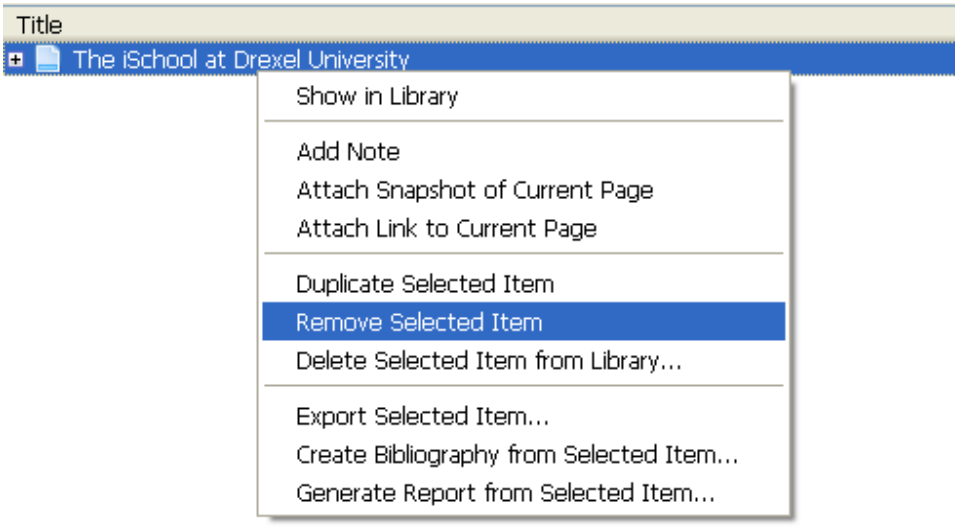


How to change a file's location?

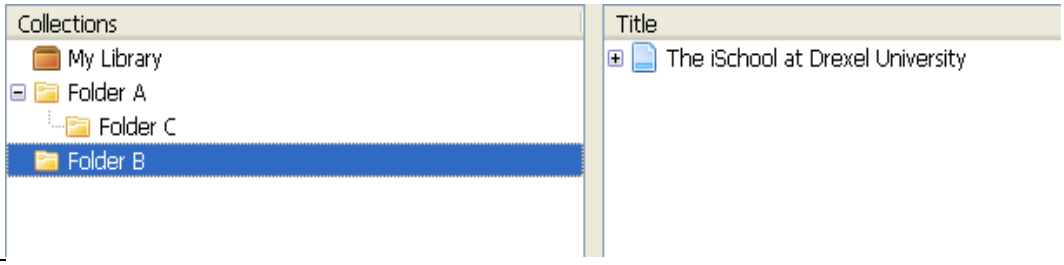
At this point, file "The iSchool at Drexel University" is saved in Folder C. Please follow the steps to move it from Folder C to Folder B.

1. Select Folder C, file "The iSchool at Drexel University" is shown in the middle pane.
2. Select the file, drag it to Folder B, and release it when Folder B is highlighted.

3. The file is now in both Folder C and Folder B.
4. Select Folder C and remove the file from Folder C. Please pay attention that you can't select "Delete Selected Item from Library…" because it will delete the file permanently from all the folders.



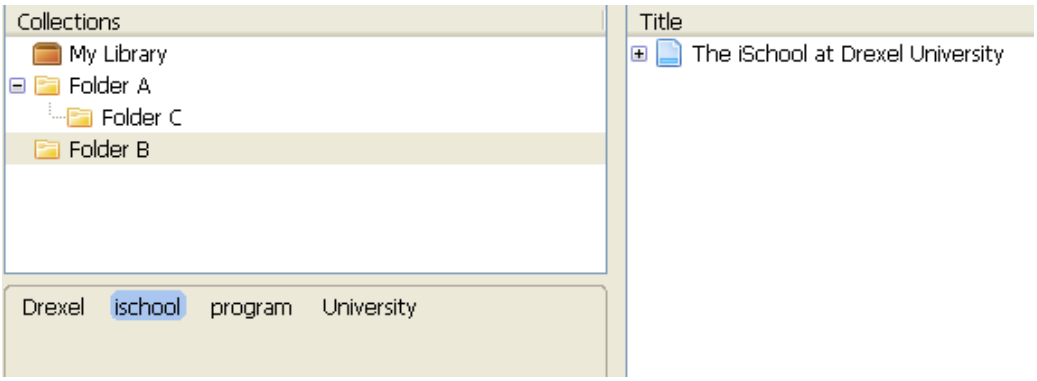
5. The file is now only shown in Folder B.



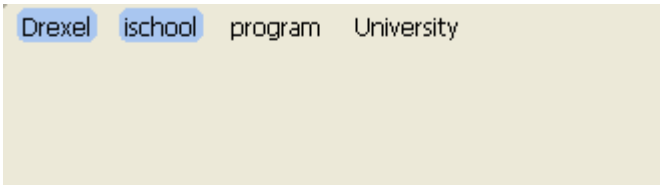
How to find your files using tags?

1. Select a target folder, in this case, Folder B. All the files in this folder will be shown in the MIDDLE pane. Tags to all files in Folder B will be shown in the LEFT pane. Tags to single file (the one you select) will be shown in the RIGHT pane.

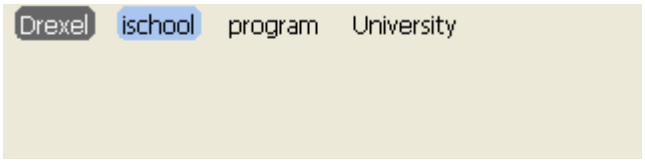
2. Click on the “ischool” tag on the LEFT pane, this tag will be shown in blue. Files that are tagged with “ischool” will be shown in the MIDDLE pane, in this case, the ischool page. All tags that are associated with ischool page will also be listed on the LEFT pane, in this case, Drexel, ischool, program, University.



3. Files can be found be by selecting multiple tags. To select more than one tag, click on the other tags that you want to select. Files that are tagged with both the two tags will be shown in the MIDDLE pane, in this case, still the ischool page we saved previously.



4. To deselect one or multiple tags, click on the blue tags you no longer want to select.

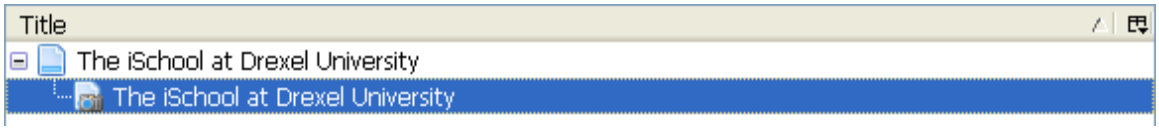


5. To deselect all the tags, click on the “Deselect” button at the LEFT bottom of LEFT pane, as shown below. None tags will be shown blue after you deselect all.

How to open a previously saved file?

1. Select the folder that the file is saved.
2. Select the file you want to open.

3. Click on the “+” sign in front of the file title.
4. Double click on the item shown below the original title. The saved file will be shown on the upper level of the window.



A-4: Tutorial on tags

Tags are one-word descriptors that you can assign to your files to help you organize and remember them. Tags are a little like keywords, but they’re chosen by you, and they do not form a hierarchy. You can assign as many tags to a file as you like or rename or delete the tags later. Tags are case sensitive in Zotero.

A list of tags on Flickr the photo sharing website (www.flickr.com) is shown below for your reference.

Architecture

Barcelona

Beach

Berlin

Bird

Birthday

Black

Blackandwhite

California

Cameraphone

Camping

Canada

Canon

Chicago

China

Christmas

Church

Concert

House

Inida

Macro

Appendix B: Task Descriptions

B-1: Introduction

Introduction

Before the study, the researcher should do the following:

- Save previous organization scheme and renew Zotero interface.

- Open fireforx. Open the ischool page for the tutorial.

- Open Zotero.

- Hide the tag selector section (For the folder group).

- Minimize the folder section (For the tagging group).

Hello! Welcome to the experiment. The experiment has two parts. The first part will take place today and the second one will take place next week when you come back. I'll be leading you through both parts.

Please first read aloud the text of the "Informed Consent Form" that you currently have in front of you. And then you need to initial and sign the form. You only need to sign one of the two copies; the other is for you to keep.

Do you have any questions so far?

Now, please fill out the pre-task questionnaire sheet.

- Hand in the pre-task questionnaire.

(Wait until the participant finishes the questionnaire.)

In today's study, you are going to use a software tool called "Zotero". Before the study start, please go through the tutorial and get yourself familiar with the tool.

- Hand in tutorial.

The interface of Zotero is shown on the screen. Please follow the steps described in the tutorial and practice. If you have any questions during the tutorial, please let me know.

(Wait until the participants go through the tutorial.)

Are you done? Ok.

Now you are going to use Zotero to finish a task. The task description is here. Your keyboard and mouse interaction with the computer will be captured electronically.

- Hand in task description for file organization session.

Please read it while I'm getting the computer ready for the task.

- Cleans up the Zotero window

- Open the index page in Firefox.

- Open Morae recorder.

Are you ready?

- Start Recording.

Hybrid group only: (There is a helping card that might be used when something unexpected happens during the experiment.)

(Wait until the participant finish the task.)

Congratulations! Please fill out post-task questionnaire 1.

- Hand in post-task questionnaire 1.

Thanks for your participation! You will come back one week later to finish the second part of the experiment.

(One week later, the participant comes back.)

- Open Zotero to the previous organization scheme based on unique ID number.

- Open Firefox.

- Open Morae Recorder.

- Open word document answer sheet.

Thanks for coming back! Today you are going to finish the second part of the study. The task description is here. Your keyboard and mouse interaction with the computer will be captured electronically.

- Hand in task description for file re-finding session.

The tutorial is opened to the page that is relevant to today's experiment. Please take a minute to look at it.

- Put the tutorial aside on the table. (Take notes when participants need help.)

(Wait until the participant finish reading task description and tutorial.)

Are you ready? Okay.

- Start Recording.

Are you done? Ok. Please fill out post-task questionnaire 2.

- Hand in post-task questionnaire 2.

Are you done? Great! Thanks for your participation! The experiment is now finished.

- Hand in the compensation for the experiment.

B-2: File Organization

Hierarchical folder group

Sixty short IT articles are listed in the main page. Please follow the links to read these 60 articles. All the articles are talking about new technology product or technology related news, including wireless, wearable technology, entertainment, and etc.

For each of the articles, please save it after you read it. File name will be saved as the name of the page by default. Please create appropriate folder for the file when you save it. If you can't decide an appropriate folder for one article, please still save it before you move forward to the next one.

After you are done with one article, use the back button on your browser to go back to the main page, and move on to the next article.

Please put them into appropriate folders so that you can find them later. There is no limit to the number of folders you create.

Please note:

1. You can not put one article into multiple folders.
2. You can spend as much time as you want.

Tagging Group

Sixty short IT articles are listed in the main page. Please follow the links to read these 60 articles. All the articles are talking about new technology product or technology related news, including wireless, wearable technology, entertainment, and etc.

For each of the articles, please save it after you read it. The file name will be saved as the name of the page by default. Please assign appropriate tags for the article when you save it. If you can't decide an appropriate tags for one article, please still save it before you move forward to the next one.

After you are done with one article, use the back button on your browser to go back to the main page, and move on to the next article.

Please give them appropriate tags so that you can find them later. There is no limit to the numbers of tags you create. You can spend as much time as you want.

Hybrid Group

Sixty short IT articles are listed in the main page. Please follow the link to go to these 60 articles. All the articles are talking about new technology product or technology related news, including wireless, wearable technology, entertainment, and etc.

For each of the articles, please save it after you read it. The file name will be saved as the name of the page by default. Please create appropriate folder and assign appropriate tags to the document. If you can't decide an appropriate folder or appropriate tags for one article, please still save it before you move forward to the next one.

After you are done with one article, use the back button on your browser to go back to the main page, and move on to the next article.

Please put them into appropriate folders and give them appropriate tags so that you can find them later. There is no limit to the number of folders and tags you create.

Please note:

1. You can not put one article into multiple folders.
2. You can spend as much time as you want.

B-3: File Re-finding

Thanks for coming back! In this session, you are going to find some articles from the collection you saved one week ago. Different cues will be given for different task environments. You are not allowed to search during the session. Please finish the task as quickly as you can.

When you find the right article, please type in the article number into the Word document. You don't need to actually answer any of the questions. If you can't find the right one, just leave the box blank and move to the next one. Please ONLY work on one question at one time.

Please finish the tasks as quickly as you can.

Task 1: Please find articles based on article titles.

Title 1: Learning coffee machine on the horizon, could use GPS/RFID.

Title 2: LG Philips announces A4 color e-paper.

Title 3: Customers reporting problems with DirectTV's TiVO service.

Title 4: Popalive remote lets you spin iPod tunes at a distance.

Title 5: Traffic reports to derive from cellphone location data.

Task 2: Please find articles based on article summaries.

Summary 1: Is your mobile conversation toxic? You will find it out in this article.

Summary 2: In this article you will learn about a new eyewear that can change transparency and color to help people easily adjust to changing lighting.

Summary 3: In this article, you will read about the reasons of not allowing some electronic devices to be used in air travel.

Summary 4: The article talks about a device that not only tells you time, but also helps you play cards.

Summary 5: The article talks about how historical exhibitions can be made using high technology.

Task 3: There is one article relevant to each of the following question. Please find them.

Question 1: What device gives visual representation of personal preferred schedules?

Question 2: Which company designed the new intelligent household machine that sends you email or messages about your food supply?\

Question 3: What are the technologies used for tracking alcohol shipping containers?

Question 4: What technology is used to know better about school kids' daily food consumption?

Question 5: What's the new way that Motorists in Ohio use to pay the toll?

Task 4: There are more than one relevant articles to each of the following task. Please find as many as you can. You don't have to fill in all the boxes. It doesn't matter if the article numbers are not in sequence.

1. Articles about entertainment system installed in different means of transportation.
2. Articles about how technology enhances the way people do exercises, sports and outdoor activities.
3. Articles relevant to online movies or movie downloading.

Appendix C: Questionnaires

C-1: Pre-task questionnaire

Participant ID: _____

Researcher: _____

Today's date: _____

1. Name _____
2. Email address _____
3. Gender □ Male □ Female
4. Are you? □ Caucasian □ Black not Hispanic

   □ Hispanic or Latino □ Pacific Islander

   □ Native American or Alaskan □ Asian □ Other
5. Age

   □ 18 or under □ 19-25

   □ 26-35 □ 35 or over
6. What is your education level?

   □ Undergraduate □ Master

   □ PhD □ Other
7. What is your major_____
8. How many years have you been using computer?

   □ 1-5 years □ 5-10 years

   □ More than 10 years □ Other: _____
9. What operating system(s) do you use at work on your primary desktop computer?

□ Windows □ Mac

□ Unix □ Other: _____

10. My files on my computer are well organized.

□ Strongly disagree □ Disagree □ Neutral □Agree □ Strongly agree

11. I usually don't have any problems finding my files from my computer.

□ Strongly disagree □ Disagree □ Neutral □Agree □ Strongly agree

12. I use desktop search to help me find files from my computer.

□ Strongly disagree □ Disagree □ Neutral □Agree □ Strongly agree

13. I've used tags on web sites such as Flikr.com, Del.icio.us, CiteUlike.org before.

□ Strongly disagree □ Disagree □ Neutral □Agree □ Strongly agree

14. How do you usually organize your files on your computer?

15. How do you usually find your files on your computer?

C-2: Post-task questionnaire 1

Hierarchical folder group

1. I successfully finished the task.

□ Strongly disagree □ Disagree □ Neutral □Agree □ Strongly agree

2. The software tool is easy to use.

□ Strongly disagree □ Disagree □ Neutral □Agree □ Strongly agree

3. The 60 articles are easy to understand.

□ Strongly disagree □ Disagree □ Neutral □Agree □ Strongly agree

4. These 60 articles are easy to organize.

□ Strongly disagree □ Disagree □ Neutral □Agree □ Strongly agree

5. It's easy to come up with appropriate folders for the articles.

□ Strongly disagree □ Disagree □ Neutral □Agree □ Strongly agree

6. I am satisfied with my organization structure.

□ Strongly disagree □ Disagree □ Neutral □Agree □ Strongly agree

7. Please describe the strategy you use to categorize the articles.

8. What are the functions you would like the system to have while it doesn't?
9. What are the problems you experienced while organizing the articles, if any?

Tagging Group

1. I successfully finished the task.

□ Strongly disagree □ Disagree □ Neutral □Agree □ Strongly agree

2. The software tool is easy to use.

□ Strongly disagree □ Disagree □ Neutral □Agree □ Strongly agree

3. The 60 articles are easy to understand.

□ Strongly disagree □ Disagree □ Neutral □Agree □ Strongly agree

4. These 60 articles are easy to organize.

□ Strongly disagree □ Disagree □ Neutral □Agree □ Strongly agree

5. It's easy to come up with appropriate tags for the articles.

□ Strongly disagree □ Disagree □ Neutral □Agree □ Strongly agree

6. I am satisfied with my organization structure.

□ Strongly disagree □ Disagree □ Neutral □Agree □ Strongly agree

7. Please describe the strategy you use to tag the articles.
8. What are the functions you would like the system to have while it doesn't?
9. What are the problems you experienced while organizing the articles, if any?

Hybrid Group

1. I successfully finished the task.

□ Strongly disagree □ Disagree □ Neutral □Agree □ Strongly agree

2. The software tool is easy to use.

□ Strongly disagree □ Disagree □ Neutral □Agree □ Strongly agree

3. The 60 articles are easy to understand.

□ Strongly disagree □ Disagree □ Neutral □Agree □ Strongly agree

4. These 60 articles are easy to organize.

□ Strongly disagree □ Disagree □ Neutral □Agree □ Strongly agree

5. It's easy to come up with appropriate folders for the articles.

□ Strongly disagree □ Disagree □ Neutral □Agree □ Strongly agree

6. It's easy to come up with appropriate tags for the articles.

□ Strongly disagree □ Disagree □ Neutral □Agree □ Strongly agree

7. I would like to tag files before I save them into folders if possible.

□ Strongly disagree □ Disagree □ Neutral □Agree □ Strongly agree

8. I am satisfied with my organization structure.

□ Strongly disagree □ Disagree □ Neutral □Agree □ Strongly agree

9. Please describe the strategy you use to categorize the articles.

10. Please describe the strategy you use to tag the articles.

11. What are the functions you would like the system to have while it doesn't?

12. What are the problems you experienced while organizing the articles, if any?

C-3: Post-task questionnaire 2

Hierarchical folder group

1. I successfully finished the tasks.

□ Strongly disagree □ Disagree □ Neutral □Agree □ Strongly agree

2. It's easy to find relevant articles.

□ Strongly disagree □ Disagree □ Neutral □Agree □ Strongly agree

3. The finding process is similar to what I usually do to find my files on my own computer.

□ Strongly disagree □ Disagree □ Neutral □Agree □ Strongly agree

4. Folders are helpful for me to find articles.

□ Strongly disagree □ Disagree □ Neutral □Agree □ Strongly agree

5. Folders remind me things about the articles.

□ Strongly disagree □ Disagree □ Neutral □Agree □ Strongly agree

6. Please describe how you found articles in task 1.

7. Please describe how you found articles in task 2.

8. Please describe how you found articles in task 3.

9. Please describe how you found articles in task 4.
10. What are the functions you would like the system to have while it doesn't?
11. What are the problems you experienced while finding the articles, if any?

Tagging Group

1. I successfully finished the tasks.

□ Strongly disagree □ Disagree □ Neutral □Agree □ Strongly agree

2. It's easy to find relevant articles.

□ Strongly disagree □ Disagree □ Neutral □Agree □ Strongly agree

3. The finding process is similar to what I usually do to find my files on my own computer.

□ Strongly disagree □ Disagree □ Neutral □Agree □ Strongly agree

4. Tags are helpful for me to find articles.

□ Strongly disagree □ Disagree □ Neutral □Agree □ Strongly agree

5. Tags remind me things about the articles.

□ Strongly disagree □ Disagree □ Neutral □Agree □ Strongly agree

6. Please describe how you found articles in task 1.
7. Please describe how you found articles in task 2.
8. Please describe how you found articles in task 3.
9. Please describe how you found articles in task 4.
10. What are the functions you would like the system to have while it doesn't?
11. What are the problems you experienced while finding the articles, if any?

Hybrid Group

1. I successfully finished the tasks.

□ Strongly disagree □ Disagree □ Neutral □Agree □ Strongly agree

2. It's easy to find relevant articles.

□ Strongly disagree □ Disagree □ Neutral □Agree □ Strongly agree

3. The finding process is similar to what I usually do to find my files on my own computer.

□ Strongly disagree □ Disagree □ Neutral □Agree □ Strongly agree

4. Folders are helpful for me to find articles.

□ Strongly disagree □ Disagree □ Neutral □Agree □ Strongly agree

5. Folders remind me things about the articles.

□ Strongly disagree □ Disagree □ Neutral □Agree □ Strongly agree

6. Tags are helpful for me to find articles.

□ Strongly disagree □ Disagree □ Neutral □Agree □ Strongly agree

7. Tags remind me things about the articles.

□ Strongly disagree □ Disagree □ Neutral □Agree □ Strongly agree

8. Please describe how you found articles in task 1.
9. Please describe how you found articles in task 2.
10. Please describe how you found articles in task 3.
11. Please describe how you found articles in task 4.
12. What are the functions you would like the system to have while it doesn't?
13. What are the problems you experienced while finding the articles, if any?

Appendix D: Answer Sheet

Please fill in the number of the article you found.

Title 1	
Title 2	
Title 3	
Title 4	
Title 5	

Summary 1	
Summary 2	
Summary 3	
Summary 4	
Summary 5	

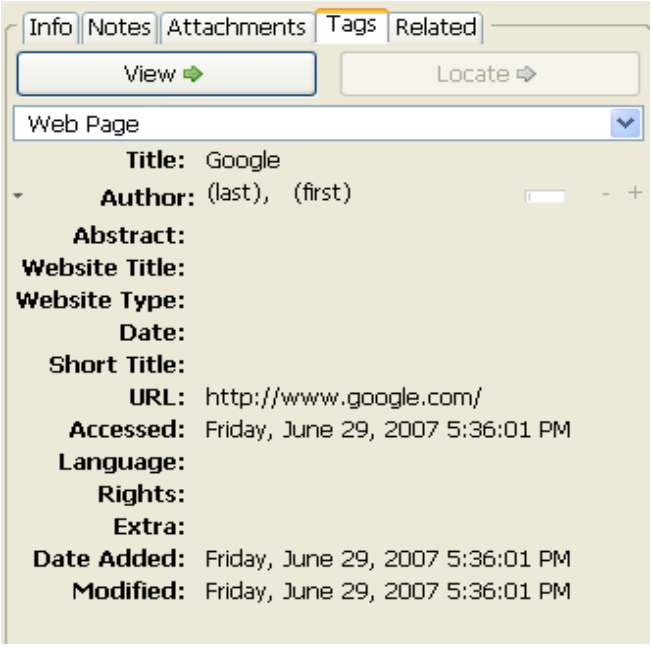
Question 1	
Question 2	
Question 3	
Question 4	
Question 5	

Question 1									
Question 2									
Question 3									

Appendix E: Help Card

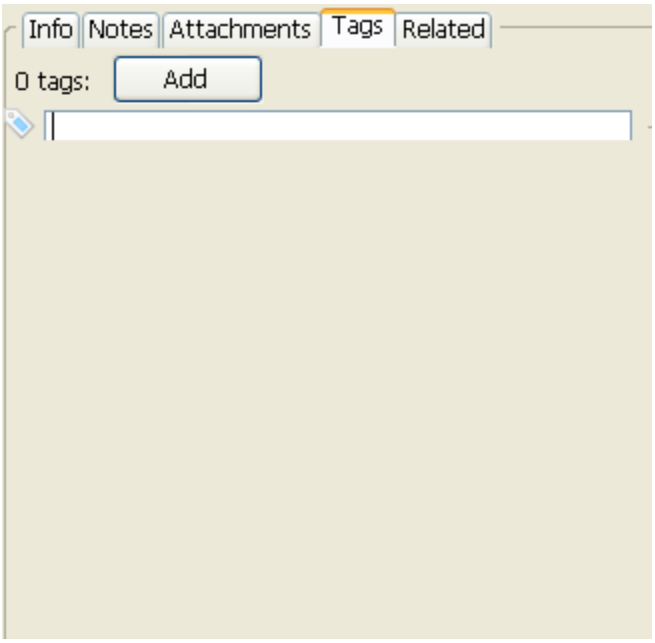
Help card for file organization session

1. If you find that you can't assign tags on the right pane as shown below.



2. Please click on any other button on the right screen such as "Info", "Notes", and then click on "tag" button.

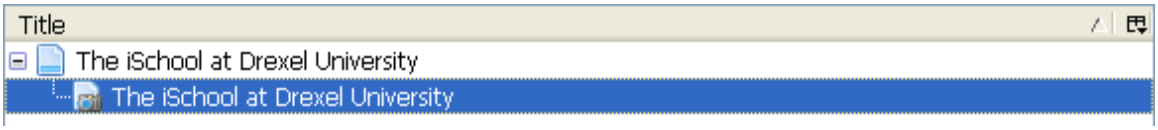
3. Now you can assign tags.



Help Card for file re-finding session

1. If files don't show up correctly, please click "Deselect All" button at the LEFT buttom of the Left Pane shown below.

2. To open a previously saved file, click the "+" sign in front of the file title and then double click the item shown below the original title.



Appendix F: The Article Collection

1 $500 million underwater fiber network to link Asia, America

by Darren Murph

Unfortunately for most, traversing back and forth to Asia from America on a regular basis isn't exactly in the cards, but thanks to a $500 million project agreed upon by a 17-member telecommunications consortium, visiting via fiber will soon be a whole lot snappier. Telekom Malaysia, along with 16 other firms, have awarded a half billion dollar contract to Alcatel-Lucent and NEC to construct a 12,428-mile link between the west coast of America and Southeast Asia. The aptly-dubbed Asia-America Gateway will connect the western US with Malaysia, Singapore, Thailand, Brunei, Hong Kong, the Philippines, Guam, and Hawaii, while also offering "seamless interconnection" with Europe, Africa, and Australia. Moreover, the project is being designed to provide a "more secure link for traffic" across the seas, as it avoids the hazardous Pacific Ring in hopes of dodging massive internet outages due to unexpected earthquakes. Best of all, the wait time for the undersea cabling to make an impact is fairly reasonable, as users should see "faster and more reliable service" when it becomes operational in December of next year.

2 Air conditioners kill hard drives?

by Evan Blass

We already knew that we had to sacrifice our precious holiday decorations in order to get maximum bandwidth and coverage from our WiFi routers, and now it looks like we may have to sweat out the summer months sans air conditioning if an anecdotal, single-source article in Associated Content proves accurate. According to sole interviewee Ben Carmichel of ESS Data Recovery, his company sees "a 20% increase in failed hard drives hitting our lab in the summer as [opposed to] the winter," and believes from surveying customers and the uptick in electronic-related issues that AC-triggered power surges are largely to blame. Of course, this alleged risk is easy enough to minimize by utilizing an uniterrupted power source for your PC and configuring your drives in a RAID 1 or 5, so it doesn't seem too difficult to have your cake and eat it too in this case. Or you could set up shop in Siberia and just avoid this nonsense altogether.

3 American Airlines offering PMPs to upper class fliers

by Paul Miller

Bored and lonely first class and business class transcontinental fliers on American Airlines will now be treated to branded Archos players to while the time away with pre-loaded video and audio programming. Such premium fliers will be able to opt for an Archos 704, while all passengers on MD-80 flights between LA and Chi-town can get an Archos 605 to play with -- though the economy peeps have to pay. AA is packing the 704 with Bose Noise Canceling headphones, and the 604 with "high-end" earbuds, and has feature films, TV, news, music videos, AA radio and music CDs pre-loaded. Right now American Airlines is running a four month test of the service, which is sure to incur a few more thefts then those headrest-mounted displays cropping up on certain airlines, but a nice bit of good will from discerning portable media consumers.

4 Apple, Hollywood go toe-to-toe on movie downloads

by Cyrus Farivar

Not surprisingly, Apple is facing heat from Hollywood companies that want to sell their movies on iTunes, but don't want to play by all of Apple's rules. The big issue is whether or not Apple will restrict the number of devices that can play a film bought from iTunes; Universal, 20th Century Fox, Paramount and Warner Bros. are rightfully concerned with rampant digital piracy and how iTunes might help inadvertently contribute to that. Of course, Hollywood could go the Wal-Mart or Amazon Unbox route if Apple won't back down, but then those studios would miss out on some of that big Apple halo action. Still, if Disney and the soon-to-be-added Lions Gate are already at the iTunes party, it seems unlikely that the rest of Hollywood could stay away much longer.

5 Artsy 8AM task lamp strikes a pose

by Darren Murph

These days, a lamp needs to be more than just a lighting instrument, and while the 8AM task lamp doesn't sport integrated speakers, an FM radio, or solar panels, it does go a long way in cleaning up clutter and reaching hard-to-light areas. Constructed from braided electric wire and an aluminum clamping system, this design-centric light allows users to create their own lighting scenario by bending and folding the post, and the rigid structure enables some fairly eye-catching poses to be struck. Furthermore, the device brightens your day more and more

each time you touch the clamp, but we're sure a clap-sensitive setup could be installed if that's your thing. Apparently, this creation is still in the prototype stage, but the commercial appeal on this one is pretty glaring.

6 AutoDocker to make boarding and deplaning faster

by Cyrus Farivar

Just about everyone has been in line while waiting to board a plane, and perhaps only the acrophobic have wished it wouldn't take so freakin' long. Well, if you're in Denver, you may just be in luck. Dewbridge Airport Systems is claiming that its new robotic passenger bridge, dubbed the "AutoDocker," (no relation to the khakis-bot) will halve boarding and deplaning time by placing walkways at both the front and rear entrances on a plane. AutoDocker has been using its sensors and 3D object recognition abilities on doors of different types of aircraft starting this week on United Airlines flights coming in or out of Denver International Airport. We're pretty sure that if AutoDocker were to combine with that new Airbus emergency landing robot we spotted earlier this year, then all we'd need now is robotic air traffic controllers to complete the mechanization of air travel. Yes, that was a joke about flight attendants being robots.

7 Automatic toilet paper dispenser doles out neater wads

by Evan Blass

No, this prototype toilet paper dispenser isn't for you lazy lumps out there, but rather serves to aid disabled folks who sometimes have trouble using those frustrating plastic boxes in public restrooms or even the simple rolls in their own homes. Called the 'TPer' -- we think -- this model operates much like a LEGO-built one we saw awhile back, spinning up an appropriate amount of neatly folded paper before slicing it free to go about its duties. There doesn't seem to be any more info on this one apart from what you see in the vid located after the break, so if you know someone who could use such a device, it looks like you'll have to break out your trusty NXT kit.

8 BMW shows off WiFi / cellular-based in-car multimedia system

by Donald Melanson

Among all the new cars unveiled at the recent Geneva motor show, BMW also took the opportunity to show off a little in-car action, letting show-goers take its still-in-development dashboard console for a spin. According to Wired blog Autopia, the system can be loaded up with audio and video content off your PC via WiFi while

you're parked in the driveway, or via a cellular link when you're on the road. Unfortunately, they didn't provide many details on the console itself, other than that it boasts 20GB of storage and will apparently be able to record radio content as well. It's also not clear when we might see such a system installed in a future BMW, with the company only saying that it "could" be commercially available within three years, adding that licensing and copyright issues are (not surprisingly) the main things holding it back.

9 Brookstone's Digital Photo Wallet eliminates need for prints

by Darren Murph

While digiframe manufacturers keep pushing for larger and larger LCD sizes to adorn (or completely consume) your coffee tables, Brookstone's going the opposite direction with its Digital Photo Wallet. The presumably chintzy "nappa leather" wallet doesn't look to be of much better quality than those sold at your favorite dollar-for-everything store, but it does manage to house a 1.4-inch LCD that reportedly stores 55 photos. Additionally, it's completely removable and rechargeable, sports an on / off button and scroll controls, and interfaces with your PC via USB. Of course, you could just stick with the wallet you're comfortable with and pick up a digital photo keychain, but if you're serious about this one here, watch for it in mid-May for $59.

10 California Senate attempts to ban RFID tracking of students

by Donald Melanson

It looks like the backlash from one California school's attempt at mandatory RFID tracking of students could end up prompting a statewide ban of the practice, if the state Senate has its way. In that first instance, Brittan Elementary School in Sutter, California attempted to require all students to carry RFID-equipped ID cards that would allow them to be tracked throughout the school, supposedly to simplify attendance-taking and reduce vandalism. That plan quickly backfired, however, and the school put the kibosh on the program. Under this new law, all schools would be prohibited from requiring students to carry RFID cards (or, presumably, be implanted with 'em) until 2011, when the practice would be reconsidered, according to The Register. Democrat Joe Simitian (who introduced the legislation) doesn't seem to be stopping there though, also proposing bills that would place a temporary ban on RFIDs in driver's licenses, add additional privacy safeguards to RFID-enabled government IDs and, of course, restrict forced RFID implants in people.

11 Canadian museum to feature motion sensitive displays, personal digital keys

by Conrad Quilty-Harper

A Canadian Human Rights Museum due to open within the next four years is to include a range of technology not usually suited to museums, including a motion sensitive display wall and a digital "human rights key" which documents the visitor's trip. The details on the technology are a little sparse, although the CanWest News Service likened the motion sensitive display wall to controlling a Wii. As there doesn't appear to be any control device involved, we'd say the implementation will be closer to that of the Precrime program, but as the building won't be ready for another four years -- by which time there'll probably be a whole new console technology for museums to emulate -- we'll just have to wait and see. The "human rights key" is equally vague, and could be as simple as a USB key that downloads information from the exhibits. Not that we've got anything against human rights, but for the sake of the kids visiting the museum, we'd hope that the key is a little more stacked out than just a USB key. Then again, if Engadget ever drafted a piece of human rights legislation, several sections would be put towards appeasing the robot population, so what do we know?

12 Casio Magic Watch

by Donald Melanson

Just when we were beginning to think that the well of watch innovation had run dry, Casio comes along and renews our faith in multi-use timepieces, with a watch that not only tells time, but does magic! The appropriately-named Casio Magic Watch was created with the assistance of close-up magician Tomohiro Maeda, who may or may not be responsible for the totally 80s-inspired design, but most definitely is the man behind the five tricks the watch can help you perform, including guessing the number or card that a person is

thinking of or roping your friends into a little con game. For some old school magic, each watch also comes bundled with an exclusive set of Tomohiro Maeda playing cards from Tally-Ho, though you'll have to decide if that's enough to justify the $80 price tag.

13 Cellphones are dangerous/not dangerous, bee-friendly edition

by Darren Murph

As the saga continues, we've got yet another flip-flopped story rolling through in regard to the toxicity (or not) of cellphones to our environment. Just under a fortnight ago, a report based on an (admittedly lacking) research study claimed that Colony Collapse Disorder within bees was being encouraged by cellphone radiation. As expected, the researchers began living a life filled with Q&A sessions about the data, and now the "truth" is coming out. Essentially, the scientists are claiming that their data was "misinterpreted," and that the study actually looked at DECT phones and base stations, which transmit a "different frequency than mobiles." Furthermore, another member chimed in and boldly stated that their "studies cannot indicate that electromagnetic radiation is a cause of CCD." So that settles it -- until the next round of bickering begins, of course.

14 Cellphones on planes banned for crowd control issues?

by Donald Melanson

While interference has long been the official explanation for why cellphones have been off-limits on airplanes, that hasn't stopped folks from speculating about other possible motivations behind the ban, with the latest bit of rumination on the subject coming to us from Mike Elgan of Computerworld. According to Elgan, one of the "real reasons" cellphones aren't allowed on flights is because of the airlines' fear of crowd control problems should phones be permitted, including the possibility that disputes could erupt among passengers as a result of rude behavior (a pretty safe bet). He also suggests that the airlines prefer to have passengers "ignorant" about problems on the ground during flights, although that issue doesn't seem to have stopped airlines from providing their own phones. Among other reasons, Elgan claims that the government wants to keep the ban in order to avoid the expense of having to test and certify every gadget for use on planes. Of course, there's also the possibility that cellphones on planes are just a really bad idea, but unfortunately we're not the ones that make these decisions.

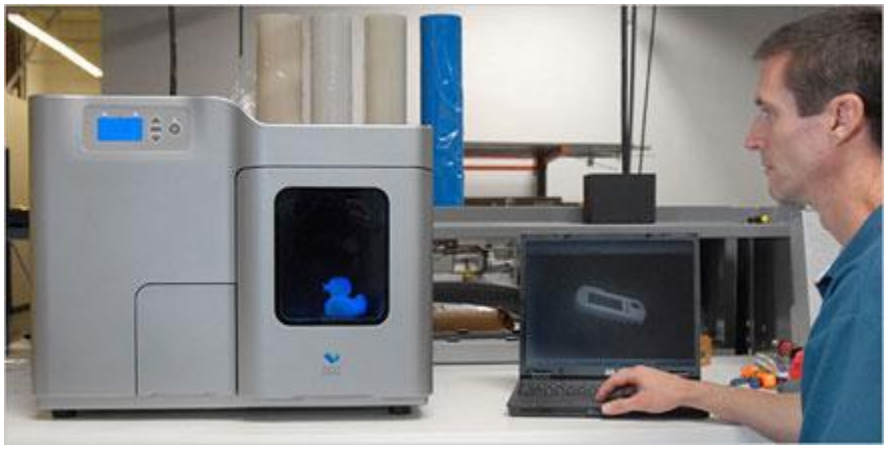
15 Customers reporting problems with DirecTV's TiVo service

by Donald Melanson

DirecTV and TiVo have certainly had their share of problems with each other, but it looks like the relationship has now taken a turn for the worse on the technical level, with ZD Net reporting that customers are encountering widespread problems with their DirecTV with TiVo service. The main complaint appears to center around the Season Pass feature, with numerous customers on DirecTV's forums posting of shows only occasionally being recorded or not being recorded at all, with some also reporting that the problem seems to be getting progressively worse. Those who complained to DirecTV's customer service about the problem say they were told that the company is aware of the issue and that it could take anywhere from one to 30 days to fix it. To appease those missing out on their favorite shows, DirecTV's apparently been offering dissatisfied customers a range of incentives to keep 'em from jumping ship, including discounts off DirecTV with TiVo service and free Showtime for a month. The latest word on the forums, however, is that the problem still appears to be unresolved.

16 Desktop Factory to offer up $5,000 3D printer

by Donald Melanson



While they're still a long ways from sharing shelf space with cut-rate inkjet printers, it looks like three dimensional printers are slowly inching towards the consumer space, with Idealab company Desktop Factory set to sell its first 3D printer for not entirely unreasonable price of $4,995 sometime this year. According to The New York Times, some 200 customers have already signed up to buy the printer, which will make up the entirety of the initial test run. They may be feeling a bit a buyer's remorse before too long, however, as the company says the price of the printers will come down to $1,000 in four years. Eventually, Idealab foresees companies selling 3D designs of products on the Internet, which people could then simply print out at home. For now, however, the printers would seem to have somewhat limited applications, relying on nylon mixed with aluminum and glass that results in gray objects that have somewhat jagged edges and a sandy finish.

17 Drager's new Infinity TeleSmart WiFi-based patient monitors

by Paul Miller

While we confess to spending very little of our free time dwelling on biotelemetry solutions, we're still pretty stoked to see WiFi cropping up in Drager Medical's new Infinity TeleSmart line of patient monitors. The new full-size monitors, suitable for children and adults, include built-in color displays to show off those vitals loud and proud, rechargeable batteries, and built-in alarms that can be controlled from the unit. Of course, the big news is WiFi. The inclusion of that 802.11 goodness makes it one of the first telemetry monitoring solutions to

support an industry-standard wireless protocol, and should ease implementation in most hospitals. And luckily, the device can continue to monitor patient data even when out of WiFi range, so there's no worry there. Hopefully they can dole some of this WiFi love for our robot friends as well, which could come in handy in those inevitable "I've fallen and I can't get up" situations. Currently the system is pending FDA approval, but we should be seeing it in the US soon after.

18 Electrochromic sunglasses change color on demand

by Darren Murph

We'll admit, we didn't think MyDo's Bururu vibrating spectacles would be topped anytime soon, and while the University of Washington's smart sunglasses don't snatch the crown outright, they certainly make a solid case. The admittedly gaudy prototype certainly isn't the most attractive eyewear we've ever seen, but researchers are insisting that we focus on the electrochromic material that enables on demand color / tint changing rather than the bulky frames. The film can purportedly alter its transparency depending on the electric current that flows through it, essentially giving athletes and motorists an easy way to instantly adjust the level of reflection and tint depending on the ever-changing circumstances around them. Touted as "more active and more intelligent" that existing specs, this watch-battery-powered device allows the wearer to alter the shade by simply spinning a wee dial on the arm, and while we wish we could tell you that U-Dub's swank new specs would be arriving before the bright summer rays, it'll be quite "a few years" before these hit commercialization.

19 EU sets short-range wireless standard for whole of Europe

by Donald Melanson

While it may still not have achieved its goal of a single currency standard for the whole of Europe, it looks like the EU may have better luck getting its member countries to make wireless products that play nice with one another, with the AP reporting that the European Commission has established standard Europe-wide frequencies for short-range wireless products (including everything from cordless phones to medical implants). Apart from the added convenience, the Commission says the move will lead to increased demand for products and lower costs for manufacturers, in theory resulting in cheaper products for consumers. Also covered under the new standards is the wireless technology everybody loves to hate: RFID, which will now be able to be even more mass produced for your tracking (or hacking) pleasure.

20 Heineken to track shipping containers via GPS

by Cyrus Farivar

We're not exactly sure why Heineken needs GPS to monitor the location of its beer shipments, but they probably know more about global shipping logistics than we do. According to RFID Journal, starting this month, the Dutch beer giant will begin the first test of its "Living Beer Plan," tracking 10 shipping containers of beer from Heineken headquarters in The Netherlands to the United Kingdom and the United States. Customs data,

such as whether or not the containers' doors have been opened, will be available online to British and American customs officials, (as well as the researchers monitoring the project at the Vrije Universiteit Amsterdam) even before the cargo ships reach UK and American ports. Heineken claims it will save on the 30-odd pages worth of printed customs forms and other documents by using GPS and online tracking instead. Surely with the money saved on not printing up documents, Heineken wouldn't mind sending its friends at Engadget a free case of brew? Pretty please?

21 Horizon vending machines set to track students' snack habits

by Donald Melanson

We've already seen a couple of systems that let parents pay for their kids' school lunches in advance and track exactly what they're eating, but it looks like they could soon be getting an even deeper glimpse into their kids' daily consumption, with the same technology now set to find its way into vending machines. Coming courtesy of Horizon Software, the snoopy new vending machines are based on the MealpayPlus system (which Horizon is also responsible for) and employ either a card or PIN code to dole out the snacks, eliminating the need for students to carry cash around. While the vending machines are apparently ready to go, it's not clear how many schools, if any, have signed up to put 'em into service.

22 Inventor crafts GPS-equipped shoes, includes a panic button

by Darren Murph

We've seen some fairly interesting means of keeping track of your mischievous kids (or pets), but Sayo Isaac Daniel's latest invention takes top honors as the ultimate paranoid parent's must-have gizmo. Aside from the obvious tracking uses, Daniel's GPS footwear is actually intended to beam out a distress signal to a pre-selected recipient if the wearer hits a certain panic button. The GPS-equipped kicks would present the location of the violated victim to whoever is deemed that person's hero, and would hopefully give the rescuer enough time to arrive and lay down the law. Also, the patent explains an "alarm toe switch" that would be inserted within the shoe in order to give customers the ability to sound their alarm (intentionally or otherwise) without making any sudden movements. Reportedly, a company dubbed Quantum Satellite Technology plans to start selling the shoes "in March for around $350 per pair," but the GPS signal emanating from your soles won't do you much good if your kidnapper ditches your footwear before tossing you in the trunk.

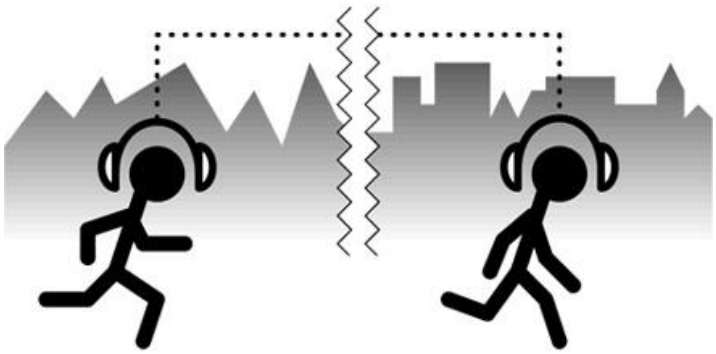
23 Inventor patents personal TV censor

by Donald Melanson

Inventor Matthew Jarman looks to be out to make the untamed TV landscape a little more palatable for those with sensitive ears, developing an application that'll mute out offensive language based on your potty-mouth tolerance level, or block programs all together if they prove to be unmutable. The system apparently works by monitoring the closed-captioning text that accompanies most television programs, muting the audio whenever it comes across a word you've deemed unsuitable -- seemingly a tricky thing to time properly, especially given the delay usually associated with CC feeds, although Jarman appears to be quite confident in its efficacy. To block entire programs, the system simply relies on program descriptions provided by electronic program guides. While there's no indication when or if the system will actually make it to market, a quick search of the ever-entertaining patent database reveals that some of Jarman's other content-blocking inventions have been licensed by none other than ClearPlay, Inc, which would seem to increase the likelihood that this one may follow suit.

24 Jogging over Distance technology makes it a social experience

by Conrad Quilty-Harper



Technology is inexorably tied to progress, but when it busts into our lives and subtly tells us that we're doing something wrong -- in this case, getting some "us" time on a solo jog -- we've gotta admit to being a little peeved. Even for those who like to share their second most sweaty experience on the day-to-day, this is probably a bit of an overkill. A prototype technology called Jogging over Distance is being developed at the University of Melbourne which uses GPS, cellphone tech, and 3D audio simulation to make it seem as if two joggers from separate cities are actually slogging it out together. The advantage over just whipping out a cellphone is that as one runner goes faster, their voice appears to be coming from further ahead to the other

runner, which could presumably help motivation. To each their own we suppose, but we'd take some loud music over the grunts of a fellow jogger any day. Or, indeed, a form of motivation that helps us -- no, forces us -- to start jogging in the first place.

25 Learning coffee machine on the horizon, could use GPS / RFID

by Darren Murph

Although a coffee machine that slowly but surely learns your daily preferences in regard to cups of java may sound outlandish, the already-created RFID-enabled refrigerator certainly brings things back into focus. A "provisional patent exploration into coffee machines that learn and react to their users" is underway in Lafayette, Indiana, as James Pappas is hoping to take ubiquitous computing to the next level on coffee makers of the future. While internet-connected and weather-displaying renditions are already on store shelves, Pappas is hoping to utilize some form of GPS / RFID technology to create a machine that learns and adapts to your coffee drinking ways so it can automatically have a white chocolate cappuccino ready and waiting each weekday (except Monday, which is your straight-up black coffee day, right?) without you having to touch a thing. Furthermore, he's hoping to take the idea to the mobile front, as he refers to a cellphone interface to dial-in your next request so that it's ready to go by the time you hit the kitchen. Still, it sounds like the invention is a few years off at best, but serious drinkers better hope this thing automatically alerts you when the beans are running low, too.

26 Lenovo's Olympics-inspired laptops revealed

by Darren Murph

Yeah, you may have thought those Santa Rosa, er, Centrino Pro-based ThinkPads were hot stuff, but they ain't got nuthin' on the future eBay value of the svelte new Olympic-inspired iteration. Shortly after Lenovo unveiled its 2008 Beijing Olympic torch design, the firm has already thrown down snapshots of its Olympic laptop, which certainly looks dashing in its own right. The stunning red and black case design is accented with a motif similar to that found on the torch, and the spiral designs actually seep beyond the external casing right onto the palm rests and around the LCD's bezel. Unfortunately, not much is know about the internal hardware that these presumably finely trained machines will sport, but early details suggests a 12.1-inch XGA display and Intel's L2400 Core 2 Duo processor will be in the mix. Additionally, mum's the word on pricing and availability, but you can rest assured that this bad boy will be one hot commodity whenever it lands in limited quantities.

27 LG Philips announces A4 color e-paper

by Darren Murph

While this doesn't mark the first time that we've caught wind of colorized electronic paper, South Korea's LG Philips has announced that an A4-sized rendition of the vivid bendable display has successfully been developed in its labs. The panel reportedly measures just 35.9-centimeters diagonally, is 0.3-millimeter thick, and can display up to 4,096 colors while maintaining the energy efficient qualities that inevitably come with using energy only when the image changes. Unsurprisingly, the company plans on marketing the device as one of convenience and doesn't hesitate to tout its greenness in the process, but unfortunately, it failed to mention when this would find its way out into the general public.

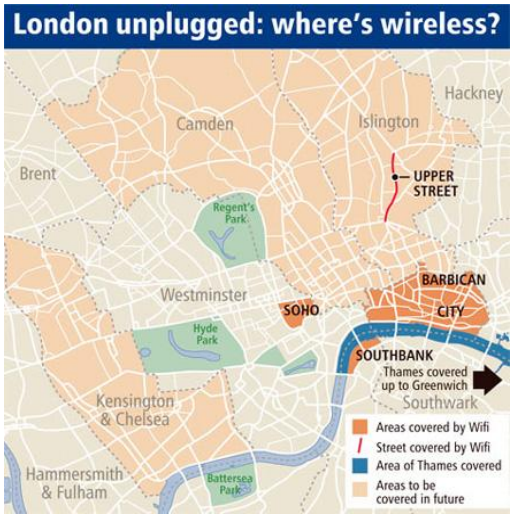
28 Logitech's new and sexy Wireless DJ Music System

by Paul Miller

There are plenty of wireless music systems already competing for our home entertainment dollar, but this new Wireless DJ Music System from Logitech is looking snazz enough to at least give a second glance. Taking a welcome cue from our usual favorite, Sonos, Logitech's remote features a clickable scroll wheel, and while their LCD isn't near as impressive, it should get the job done. The system's StreamPoint software can hook up with popular jukeboxes like iTunes, Windows Media Player and Musicmatch, and uses some 2.4GHz wireless tech for digitally streaming your tunes to the included Music Receiver, which also happens to be a dock for charging up your remote. Logitech has the price set fairly low at $250, but there's no word on when this setup is due to hit the streets.

29 London to become Europe's largest WiFi hotspot

by Darren Murph



As if blanketing the rim of the River Thames with WiFi wasn't good enough to get us out of the office and into the park, the Evening Standard is reporting that the whole city of London will soon becomes "Europe's biggest wireless internet hotspot." As expected, some 130 base stations will be arranged in a sophisticated mesh networking setup, which will span "the entire Square Mile," subsequently giving about 350,000 employees in the area access to unadulterated wireless internet. Wireless gurus from The Cloud are working in conjunction with city officials to tie off the final steps, and while initial coverage areas will dwarf Soho and Barbican City, the map above shows just how broad the service could get. Unfortunately, this edition of citywide WiFi will not

come gratis, as users who plan to take advantage will be kindly asked to fork over about £11 ($22) a month for access.

30 MasterCard tries to best EZ-Pass with PayPass

by Omar McFarlane

While riding around the roads, paying for tolls have always either been cash or a windshield mounted transponder like EZ-Pass or FasTrak. Well, for some motorists in Ohio, there's now a third option. For the next three months, select exit tolls along the Ohio Turnpike will be outfitted with self-service machines that accept MasterCard's PayPass. The trial run allows drivers wielding the appropriate cards to easily pay their toll with the contact-less system, not unlike the EZ-Pass. Vending machines around those exits will also sport PayPass options for added convenience. Hopefully, this convenience won't also be making life easier for hackers.

31 Mega hands-on: Virgin America's Airbus A320 with Red in-flight entertainment

by Ryan Block

Not-yet-airborne Virgin America invited us to check out the way-decked Airbus A320 with Red prototype in-flight entertainment system that's parked at SFO right now. (Naw, we didn't get to take it up, the US Dept. of

Transportation hasn't yet cleared VA for commercial flights yet, boo.) They definitely weren't kidding when they said it's got it all: movies on demand, pervasive music playlists, in-seat messaging with a QWERTY controller, touchscreen Linux consoles with games, the works. We've got a massive, massive gallery for you to check out (it's really not to be missed); we toured the aft cargo area where the each plane's servers live, the Red in-seat consoles (of course), the cockpit, even WiFi-enabled flight attendant handhelds. We've also got a full rundown of everything you need to know about Virgin America, Red, and the kitted-out Airbus, so check out the gallery below, and click on for more details about the only airline we officially sanction as being geeked enough to transport Engadget.

32 MGM movies hit iTunes

by Donald Melanson

It's a little late to the party, but MGM has finally joined Paramount, Lionsgate, and others in offering a selection of its films for download in the iTunes Store. It's not the first time MGM has served up some of its movies in digital form though, with the studio already having inked a deal with Amazon's competing Unbox service. Those that prefer not to stray too far from iTunes, however, can now snag titles such as Ronin, Dances With Wolves, Mad Max, Rocky, Foxy Brown and, yes, Robocop at the usual "near-DVD quality" resolution for about ten bucks a pop.

33 Microsoft and Yahoo: no deal

by Ryan Block

According to The Journal, the Microsoft / Yahoo merger talks which first took place a year ago (and then again "in recent months", as was discovered today) turned up nil -- that's right, the deal's off. Again. Not that it was ever on. Anyway, apparently discussions are in fact no longer active, and it was concluded that competing is apparently marginally more productive than merging for the two mammoth businesses. (Seriously, can you imagine how much product overlap there would have been between the two?) And trust us, the last thing 75k-large Microsoft needs right now is another personnel infusion of over ten thousand people having a hard enough time as it is keeping up with the business. Bottom line: looks like both companies are on their own against the big G. Best of luck to 'em.

34 My Voice Remote gives your fingers a break

by Cyrus Farivar

So if you got that One Voice Media Center Communicator when it came out last year, either you fell in love and remained parked in front of your media center all the time, or you gave up on it because you found that sitting so close to your flatscreen was killing your eyes. One Voice must have gotten a real deluge of emails from weary-eyed customers, because the company has just combined the simplicity of a time-tested television

tradition, the remote control, and merged it with a microphone -- calling it the My Voice Remote. That way, instead of making your neighbors think that you've gone totally insane by barking commands at your media center, you now can order it around by talking into your remote directly from the comfort of your lounge chair. As if that wasn't weird enough, you can also use the remote as a Skype handset (seriously), which isn't quite at the same level of ridiculous-but-still-cool as Maxwell Smart's shoe phone, but we're sure that's coming next.

35 NanoNuno: the water-repelling umbrella

by Darren Murph

So we've got umbrellas that are WiFi-enabled, iPod controllers, and wannabe meteorologists, but how about one that just keeps itself (and you) dry? Some crafty blokes in Britain are getting back to basics with the NanoNuno umbrella, which utilizes nanotechnology to create a canopy of minuscule fibers that water, dirt, and a host of other elements out there have a hard time adhering to. The idea was supposedly borrowed from the "natural phenomenon" witnessed on the Lotus leaf, which sports a similar, sleek surface that water simply rolls off. Since moisture can (presumably) not penetrate the surface, the company claims a simple shake or two will return the umbrella to the arid state it was in before braving the weather, but priced at £49.95 ($94), this probably only appeals to those stuck on the windward side of the island (or Seattle).

36 Netflix founder joins Microsoft's board

by Thomas Ricker

Get ready to mop the juices fanboy, Microsoft just announced that Reed Hastings, chairman and chief executive of Netflix, will join their 10-member board of directors. So Microsoft -- the people behind the Xbox Live video-download service -- and Netflix -- recently seen pushing into digital delivery of video -- are now officially in cahoots. Of course, we must be careful about reading too much into board-level appointments. After all, no business deals were included in the announcement. But damn if this doesn't sound like a potential win-win for both camps: Microsoft could quickly extend their meager video catalog while Netflix finally lands themselves a living room box. With both facing stiff competition from the likes of Amazon/TiVo, Wal-Mart, and now Apple, it sure smacks of something in-the-works, eh? Oh, and it wouldn't be too bad for us consumers either.

37 Netflix "Watch Now" feature offers movie download alternative

by Paul Miller

Just as online music stores introduced the "rental" model to music listeners, it looks like we're finally getting beyond the rent / purchase model for online movies, with Netflix mixing things up with its new Watch Now

feature. A user gets an hour of video watching for every dollar they spend monthly on the service -- a $17.99 subscriber gets 18 free hours of video. There's no concept of buying or renting a movie, instead you just watch what you want to watch, with every minute spent watching each movie counting against your total, but nothing more. This helps users catch the last few minutes of recent rentals they might not have finished, or preview a bit -- or the whole thing -- of a prospective selection. With a good enough connection, the quality apparently approaches DVD, but there aren't any portability options or DVD burning here: it's all browser based and not a whole lot more. The feature is also XP only at the moment, but Netflix hopes to expand to other platforms, including cellphones. Currently there are roughly 1,000 titles available, which Netflix hopes to expand to 5,000 this year. Netflix is live with Watch Now for select customers right now, and will slowly expand to the rest of its subscribers over the next six months. It seems we've been waiting for Netflix to get into this gig for years -- oh wait, we have -- and while it does seem to be taking a bit different angle to the biz, we're glad to have 'em all the same.

38 New York City to get Taxi 2.0, now with GPS tracking and TV

by Cyrus Farivar

Cabs are to New York as freeways are to Los Angeles; that is to say, they're an integral part of the landscape and no matter how hard you try, you're going to have to engage with them at some point. As such, most New Yorkers have a love-hate relationship with taxis. But New York's Taxi & Limousine Commission is trying to ease that relationship a little bit -- on Thursday, it gave a preview of the next generation of the city's yellow fleet, where new cabs will have a touch-screen that lets you watch TV (or turn it off), pay by credit card, follow your progress on a map, and best of all, make it easier to retrieve that umbrella that you forgot in the backseat when you got dropped of that one drunken night in the East Village. The commission said that the new system will let

you call a hotline and tell them where you were dropped off and what was lost to use as clues to retrieve your forgotten property. We're assuming that then they'll retrace various cabs' driving histories (via GPS) over the period of the last several hours. But just keep in mind that you probably weren't the only person getting out at Port Authority around noon last Wednesday. The TLC says the new cabs should start hitting Manhattan by the end of the year.

39 Nike plans to make all its running shoes iPod-compatible

by Paul Miller

Not a lot of details just yet, but it looks like that limited selection of Nike+ shoes won't be alone in iPod compatibility in Nike's running lineup for long. Nike President and CEO Mark Parker said in a recent interview that Nike "plans to make all its running shoes compatible with its Nike+ technology by the end of the year." That's good news for the iPod kids looking for a wider selection of running shoes, and unwilling to go with one of those hacks, but we suppose a purist or three isn't going to like having their running shoe cut up in such a manner. Parker continued, saying "I won't get into specific dates or executions, but you can expect to see some of these new Nike concepts at retail this calendar year." Our guess is those shoes could be hitting 'round the time Nike Speed+ and friends start making this whole distance tracking thing ubiquitous in May.

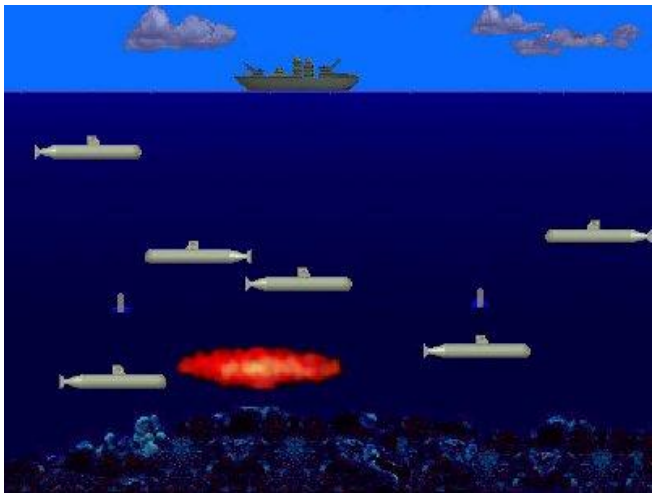
40 NYC cab drivers say "no thanks" to GPS installation

by Jeannie Choe

As the Bloomberg administration sets out to install a high-tech GPS video monitor in every NYC taxi cab, many drivers are prepping to strike. Philadelphia cab drivers faced the same challenges last year and decided to strike, however, units were installed anyway -- indicative of what we might see later on this year. The monitors would be mandatorily installed in the backs of all 13,000 NYC cabs, providing entertainment with commercials and allowing credit card transactions -- great for the customer, but what do drivers get? Well, they'll pretty much get the shaft. The meter activates GPS tracking of each trip's beginning and end destination, which many drivers claim is an invasion of privacy. Also, drivers who are buying their vehicles will be forced to shell out $2,900 to $7,200 for use of the units over a three-year period. The real kicker is that they totally dropped the ball on any sort of navigational support. Hello? Spending major cash on new taxi-gadgets with GPS technology should, first and foremost, mean we no longer have to tell clueless drivers how to get around the city. At least we'll have the "entertainment with commercials" to distract us while taking the scenic route.

41 Office of Naval Research patents "underwater GPS"

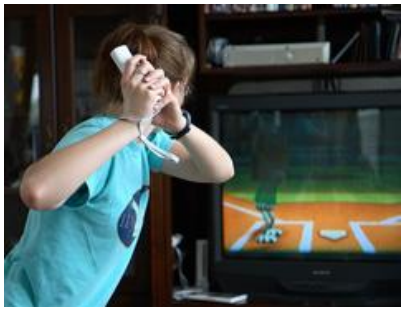
by Evan Blass



Being decidely land-dwelling creatures -- our only real underwater experiences were some nasty childhood dunking incidents at the local pool -- we've got no real use for determining our location beneath the waves, but apparently the folks who drive around all day in submarines do, so the US Office of Naval Research has whipped up a relatively simple method for "geophysical positioning" when satellite GPS data is inaccessible. You'll recall that we recently saw another patent which claimed to offer the same functionality for above ground, indoor locations, but that involved using surface-level receivers and ceiling-mounted repeaters -- a setup not possible when satellite signals can't penetrate deep underwater. The ONR's solution is to anchor an undetermined number of transceivers -- whose exact spatial location is already known -- to the ocean floor, and then by comparing the angle and distance of broadband acoustic signals sent back and forth to moving craft, using some basic geometry (sounds like our man Pythagoras comes into play here) to determine where in the world that sub happens to be. Seems pretty obvious when you think about it, but maybe that's we're covering this tech secondhand instead of trying to hack it as naval engineers, marine cartographers, or whoever the heck it is that comes up with this stuff.

42 Personal trainer uses Wii for workouts

by Conrad Quilty-Harper



As anyone who's managed to get their hands on a Wii will no doubt have already discovered, it's relatively easy to work up a sweat and lose some pounds using Nintendo's little white box. Now, the real life activity that's required to succeed at virtual Wii Sports has been picked up on by a personal trainer from Glasgow, Scotland: Zander Urquhart found that kids and adults alike don't mind doing exercise when it's accompanied by on-screen actions by the player's Wii, and is using the Wii specifically as an exercise machine. Zander is apparently the first fitness expert to have seen the potential for the Wii in the field of personal training, but we have a strong feeling he won't be the last. Even if the idea doesn't catch on, we suspect he'll have a healthy stream of geeky customers coming into the gym as long as the Wii stock shortage continues to exist (which, judging by the sporadic shortages of the 10 month old DS Lite, could be some time yet). Be honest with us: have you lost any weight thanks to your Wii?

43 Philips introduces wireless medical tablet powered by Intel's MCA platform

by Jeannie Choe

Philips and Intel will launch a new wireless, handheld medical device later on this year. This news arrives just as Motion Computing has unveiled a competitive product, the $2,200 C5 medical tablet PC. Both Motion's and

Philips' devices are based on Intel's mobile clinical assistant platform, a system designed with the intention to improve accuracy, efficiency, and work flow for healthcare professionals. Philips' new tablet device will sport a 10.4-inch, 1,024 × 768 touchscreen as well as WiFi and a digital camera, aiming to streamline data entry and better monitor wounds and healing. The device will allegedly read both barcodes and RFID tags to prevent medication errors and confirm staff and patient ID; here's to popping antibiotics, not hormone pills, after surgery!

44 Popalive remote lets you spin iPod tunes at a distance

by Cyrus Farivar

If there's one downside of having an iPod dock speaker setup in your corner, it's the fact that in order to skip a track, you have to actually walk up to the iPod and press the forward button. Fortunately, various companies have been hard at work on solving this problem by making additional peripherals for existing iPod accessories. While KeySpan already came out with one of the first iPod remotes, called the TuneView, complete with a screen and controls, Alive Style is ready to release a rival remote -- the Popalive -- which looks surprisingly like a knockoff iPod. But possibly the best part of the Popalive is that it uses radio waves, so you can control your iPod through walls just to freak your housemate / spouse out -- that is, if you're willing to throw down $150 for the pleasure.

45 'Reverse Alarm Clock' keeps the kids in bed so you can party

by Evan Blass

Three cheers for Professor John Zimmerman, who's finally doing some research to benefit Joe Public, and who has invented a device that lets parents sleep in late and put the kids to bed early on those long summer days. Zimmerman, of Carnegie Mellon's School of Design and Human-Computer Interaction Institute (and probably a parent himself), designed the so-called 'Reverse Alarm Clock' to give the tykes a visual representation of their expected schedule; when the clock's 'Sky Display' shows a sun, young children know they're free to roam about the house, but when a moon and stars appear, they'd better not get out of bed lest the boogie monster devour them whole. So far the system -- which uses a parent-set sunrise and moonset calculator, and also features a "Treasure Chest Music Selector" to pleasantly awaken your own little treasures -- doesn't seem to be commercially available, which is kind of a shame, because we know more than a few people who would love to trick their tiny terrors into bedtime at five o'clock on the daily.

46 RFID driving licences could be used for Canada-US border crossings

by Conrad Quilty-Harper

Washington State drivers are soon to be subject to a pilot trial of RFID embedded driving licenses that will be used to facilitate the passage of Americans returning from trips to Canada. The new licenses will be issued from

January 2008 and will look almost identical to traditional licenses, but will include information about citizenship and other unnamed variables. For now, the new cards will remain optional and will cost only $40 (compared to a $97 passport) but will require an interview in person along with proof of American citizenship. The underlying reasoning for the trial is the changes to federal security rules which make it a requirement for all citizens -- American or not -- to prove their identity when entering the US by sea or land. Accompanying the trial is the usual series of justifications by authorities like Homeland Security Secretary Michael Chertoff who said "It is security for our nation and our state. But it's also security for the individual"; Stockwell Day, Canada's public safety minister, even threw in the obligatory 9/11 reference by reminding us that "Canadians also died in those towers". Oh, where were the insecure RFID tags on September 10?

47 San Francisco Bay Area Bus-Fi scheme to dish out free, mobile internet

by Conrad Quilty-Harper

Public transport in the San Francisco Bay Area is to join the 21st Century now that AC Transit, one of the transit authorities in the region, is about to start testing WiFi on a fleet of 79 buses. The service will be free for riders (and freeloading WiFi addicts in following cars), and is due for a full rollout in mid-fall. The bus routes taking part in the test phase travel along three of the longest bridges in the area (the Dumbarton, San Mateo and Bay), which is intentional: having WiFi on longer routes makes it easier for passengers to justify getting out and booting up their laptop. The idea behind the scheme is to offer a competitive advantage for buses over other forms of transportation, but there are still a few questions about the concept of WiFi-enabled buses that this test may be able to answer. Number one on our list is "will passengers be willing to use their laptops on buses?" There's a glaringly obvious security concern here: you're asking to be mugged if you're happy with pulling out

your prized laptop on a city bus. The other pertinent issue is priorities. For most commuters, having a reliable and comfortable service will always be more important than internet access. Finally, it appears that the program will be completely separate from the GooLink partnership that will eventually bring WiFi to the entire city. Joining up with the aforementioned scheme is certainly an avenue that the transit authority should look down, preferably before the state of California plunks down $340,000 of funding on the Bus-Fi scheme.

48 Segways banned in the Netherlands due to a "lack of brakes"

by Darren Murph

It looks like those serious about motoring around on their Segway might have to start crossing international boundaries in order to do so legally, as the Netherlands have now joined Japan and Britain in the growing list of countries which have outlawed the oft ridiculed (and potentially dangerous) personal vehicles. Dutch police have officially enjoined Segways "on all public roads, sidewalks, and bike paths," claiming that the lack of an onboard braking system prevents it from being categorized as a "vehicle" by the Royal Traffic Agency, which bars it from receiving a license plate and becoming a street-legal ride. Although a spokesman from the RTA actually commented that the Segway was "a nice vehicle," using the machines on public property is no longer permissible, but it was said that it could take some time before regulations are actually passed down and "enforced." Segway Netherlands director Piet Kruijt was (unsurprisingly) upset by the ruling, and claimed that he was "working on all fronts to get things resolved," and for nothing more than our sincere concern for the Amsterdam Segway Tours employees that are hoping to have work come March, we hope he's successful.

49 Siemens refrigerator gets hacked, adds RFID communication

by Darren Murph

Although we'd prefer at least a few things in our domicile remain non-intelligent, hooking our refrigerator up with a mind doesn't sound like a half bad idea. Sure, Samsung's already on the ball, but anyone interested in receiving SMS alerts about how out of date their milk is, which peanut butter not to buy, and whether Sally's favorite sherbet is all but empty isn't down with waiting another year. Thanks to Kim Otto of Denmark's Innovation Lab, along with RFID labels / readers, copper wiring, a PragmaSoft-enabled computer, and a (previously) TV-equipped fridge, the task has already been accomplished. Reportedly, the Siemens smart fridge prototype can judge all sorts of facts from foodstuffs that are tagged with RFID, and considering that it's also connected to the internet, it can be made to beam out emails or text messages to alert you of dwindling supplies or warn you of recalled produce. Best of all, the folks even took the time to video the newfangled creation, so be sure to click on through for the not-even-for-sale fridge of your dreams.

50 SkyKap Advisor: golf claps for GPS golf caps

by Ryan Block

GPS and golf meet yet again in SkyKap's new Advisor, a line of visors and caps with GPS built into the bill, bone conduction, audio built into the side, and speaker-independent voice recognition, all with the intent of keeping track of scores, times, yardage, and so on. Of course, all the while you're still going to look like a golf dork talking to the invisible voices in his head, but we understand no cost is too high to shave a few strokes off your game -- or psych out our hated opponent.

51 Target whines to major studios about online movie sales

by Paul Miller

If you can't beat 'em, complain about 'em. That seems to be the mantra of Wal-Mart, and now Target, both mega-retailers who seem to have given up the fight with video download retailers before it begins. The latter of the pair has come public with its objections to movie price schemes, calling them less expensive than DVDs, and busting out the vague threat that Target "would reconsider its investment in the DVD business" if the pricing didn't level out. Somehow, we can't quite muster much sympathy for the whiny retailer, given the fact that the online prices are hardly bargains, the resolution is lower, the special features are non-existent, and the market is still teensy tiny. Walt Disney Co. has apparently pointed most of this out to Target already, and since they've gone ahead with their open letter to the industry, we suppose it didn't do much good. Still, we've got a good feeling about market pressures pushing through full-blown online movie sales at decent prices within, oh, let's say the next decade or two, no matter how much "investment reconsidering" Target does in the meantime.

52 Targeted advertising coming to a commercial break near you

by Darren Murph

Although Microsoft has patented an advertising "gotcha" to insert fresh plugs into previously recorded shows, Visible World and OpenTV are taking targeted advertising to a whole new level. While the systems have been trialed for some time now, Visible World's variable ad system will hit the mainstream when Wendy's commercials on Fox Sports' NFL broadcasts feature raccoons that seemingly know what just happened on the field. The 'coons will bust out comments about how boring the scoreless match is, how crazy that touchdown pass was, or how miserable that shirtless guy in the front row must be in the below-freezing weather. Eventually, both firms hope to focus ads right down to individual households, claiming that "dog-related advertisements" would show up primarily in dog-owning households, and that Ford commercials could be specifically shown to compete with local / regional competition in a given area. Still, for those programs that we can stand to watch later, we doubt a slightly clever advertisement will keep our fingers off the FF button, but at least this stuff makes those live broadcasts a bit less painstaking.

53 The Ktrak: half-ski, half-track mountain bike

by Conrad Quilty-Harper

We know what you're thinking: extreme ski biking is cool now? Or alternatively, who the hell would get on one of these? Just take our advice, next time you find yourself off-piste, keep your eyes peeled for Ktrakers -- crazy people who've bought a universal attachment that replaces the wheels of a mountain bike with a track drive and an optional front ski. Apparently the Ktrak's insane appeal comes from its ability to let cyclists ride over previously inaccessible surfaces like snow and sand, as the company's videos demonstrate. Due for a production run in early 2007 (priced at a reasonable $400 for the track and $140 for the ski), the Ktrak is apparently very easy to install and ride, and adds only 5-pounds to the average mountain bike's weight. If you can manage to convince your brain that this thing actually exists, your biggest problem is going to be getting your hands on one before everyone else: to that end, you can reserve a kit for $300, as long as you get your pre-order out before January 1.

54 Tin foil hats finally find a purpose: Cool-Cap baby cooling

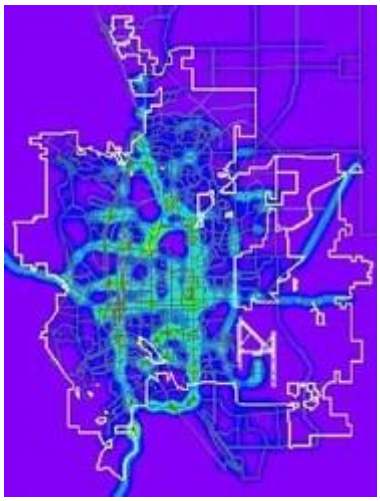
by Paul Miller

The FDA has just lent its stamp of approval to the new Olympic Cool-Cap from Olympic Medical of Seattle. The shiny little hat is designed to help treat hypoxic-ischemic encephalopathy (HIE -- yeah, we never heard of it either) in newborn infants. It's a potentially fatal injury to the brain caused by low levels of oxygen, and kills 20-25 percent of its victims, with 25 percent suffering permanent disabilities. The Cool-Cap is designed to combat the condition by creating a steady flow of water in the cap to cool the baby's head. It's not as simple as it sounds, with a whole control system involved to monitor the kid, and a solid-state water cooler to deliver uber-

precise temperatures to the cap. The foil of the cap is designed to minimize heat, which is great, since it turns out those tin foil hats of ours were never doing much good after all.

55 Traffic reports to derive from cellphone location data

by Darren Murph



Intelligent roadways and traffic monitoring systems have been available (albeit not always entirely accurate) for some time now, but if IntelliOne and AirSage have their way, finding out about real-time roadblocks (and voyeurism lawsuits) could become a more fleshed-out reality. The firms are looking to utilize that oh-so-telling "anonymous" location data from each traveler's cellphone to pinpoint locations and overlay that information with maps. If wireless companies open up that data at a rate of "twice per second" while users are conversing and "once every 30 seconds" when not on a call, the entrepreneurial duo hopes to offer more detailed information and pragmatic advice than "radar, helicopters or cameras" currently do. While keeping a keen eye on traffic developments certainly has its benefits, the real issue here is privacy (or the lack thereof); while government uses have already been in place, carriers are (understandably) more hesitant to turn over consumer data for locating purposes. While the service would be marketed free of charge to wireless carriers, interested customers not enraged by such intrusions could purchase the data for a monthly fee, and if all goes as planned, the Tampa pilot that is currently ongoing will lead to "40 other markets" being invaded by this time next year.

56 US military to launch WiFi router into space

by Darren Murph

The mysterious expanses of space have been seeing quite a lot of action lately, but rather than watching a couple of autonomous crafts tackle tasks for kicks, the US military is hoping to set their next router right beside a constellation. Partnering with Intelsat and Cisco Systems, the US Department of Defense is hoping "to test an internet router in space in hopes of benefiting civilian broadband satellite communications." Additionally, the Internet Routing In Space (IRIS) project aims to assist in all sorts of "military communications," as it routes IP traffic between space-bound sats and cuts down on the time required to divvy out information. As expected, voice, video, and data will all be supported, and once the three-year initiative is complete, "the technology will be available for commercial use" -- you know, in case you're looking to relocated your networking gear to the Black Hole.

57 Utah lawmakers condemning open WiFi networks

by Darren Murph

We've already seen where mooching off an open WiFi signal can land you in Singapore, but are we really to that point here in the US of A? Apparently, Utah lawmakers are considering "penalizing those who leave their wireless networks open" as they trial "various methods of quelling free speech controlling questionable internet content." The paranoid officials are seemingly attempting to "reward ISPs that self-police access to pornography," as the primary concern seems to stem from open signals leading to unmonitored porn surfing by kiddos under 18. Interestingly, one proponent of the plan actually goes so far as to criticize the "unregulated internet," presumably suggesting that a world of controlled, censored, and dictated material would create a much more amicable environment. Still, one (level-headed) local ISP owner stated that shutting down free WiFi zones would damage Utah's reputation, as it would appear as a locale that "is restricting technology rather than expanding it." Truthfully, we tend to agree.

58 Wal-Mart joins the digital movie download fray

by Paul Miller

Looks like all those rumors and conspiracy theories regarding a Wal-Mart download store actually counted for something this time around. Just in time for the holidays, Wal-Mart has announced its very own video download service, which will kick off in "beta" mode next week with an exclusive Superman Returns bundle. Buyers of the DVD will have an option to also obtain a portable, PC, or dual license for a movie download, priced at $1.97, $2.97, and $3.97 respectively. Users will be directed by a sticker on the DVD case to walmart.com/superman where they can enter a promo code and start the download -- which can be watched while the movie is downloading if they can't bring themselves to pop in the actual DVD. This is quite a different tact than other movie download services, such as iTunes, Amazon Unbox and Xbox Live, and it seems primarily designed to protect video sales cannibalization while leaving out much of the convenience of a traditional movie download service. Wal-Mart says it has more bundles planned for the beta launch, along with what is presumably a more traditional download service with more traditional pricing, though deets are slim at this point. What we do know is that the war for your holiday download dollar will be a bloody one this year.

59 Wal-Mart to use infrared to track shoppers / promotions

by Darren Murph

As if its insanely coordinated logistics system, biometric payment system, and (potential) RFID shelving weren't eerie enough, America's largest retailer is taking consumer voyeurism one step further with the use of infrared technology. In an apparent attempt to avoid the taboo "RFID" flavor of intrusion, Wal-Mart is hoping to sneak an IR system into its stores to gauge the effectiveness (and elicit more advertising dollars, of course) of its various promotions. Dubbed Prism, the arguably dodgy system was crafted by Coca-Cola, Kelloggs, Kroger, Procter & Gamble, Walgreens, and Disney in order to "track shoppers' movements around the store" and correlate them with actual sales in order to judge display effectiveness. The consortium of firms has coaxed the

corporate giant to install a trial system in ten of its SuperCenters, with a much broader rollout expected to follow soon; so when making that mad dash to the Tickle Me Elmo eXtreme (or bathroom supplies) section, just remember that Big Brother could be keenly watching.

60 Wi-Aquarium: the WiFi-enabled fish tank

by Darren Murph

We've seen everything from WiFi-enabled PMPs, robots, and even gardens, but a recent creation showcased at the Embedded Systems Conference in California took remote access abilities underwater. While not as profound as a bay-patrolling nuke detector, Lantronix's Wi-Aquarium was able to grab the bronze for its snazzy internet-enabled fish tank, which reportedly "allows users to remotely control and monitor their aquarium anytime from anywhere in the world." Moreover, users can login and keep an eye on the situation via webcam, and they can even dictate the water temperature, lights, and filter from afar. The standout feature, however, is the ingrained ability of the tank to send the owner emails containing status reports of several key metrics, which should certainly keep you connected to your sea-dwelling pets no matter your location. Now this is what Tamagotchi should have been from the start, eh?

Vita

Shanshan Ma worked on her Ph.D. degree in the College of Information Science and Technology at Drexel University during 2004 and 2009. She received her bachelor's degree from Sichuan University in China in the year of 2004.