Personal Document Management Strategies

Sarah Henderson
Department of Information Systems
and Operations Management
University of Auckland
Private Bag 92019

s.henderson@auckland.ac.nz

ABSTRACT

Personal document management describes the activities performed by an individual in creating, acquiring, organizing and maintaining collections of their documents. A study involving field studies and a survey of 115 participants was conducted in order to better understand the approaches people take to document management. Qualitative analysis of a field study and quantitative analysis of a survey were used together to develop a description of three major approaches to personal document management: a piling strategy, a filing strategy and a structuring strategy. A user persona was developed to exemplify each strategy; this persona description can be used as a design tool to guide the development of user interfaces for personal document management system. Specific user interface guidelines are suggested to support each of the three identified strategies.

Categories and Subject Descriptors

H.5.2 [**Information Interfaces and Presentation (e.g. HCI)**]: User Interfaces

H.3.3 [**Information Storage and Retrieval**]: Systems & Software

General Terms

Management, Design, Human Factors, Theory.

Keywords

Personal document management, personal information management, document management strategy, personas.

1. INTRODUCTION

Personal document management is the activity of managing a collection of digital documents. The unit of analysis in personal document management is an individual user and the collection of digital documents he or she owns. The process of document management incorporates the creation/acquisition, retrieval, organizing and maintenance activities described above, provided they are performed by the document owner. Personal document management is an activity that is performed intermittently, embedded in the daily life of users.

Most people store their documents in the hierarchical file system provided by their computer's operating system, and manage these

documents through a hierarchical file browser (such as Windows Explorer) [8]. These file browsers were intended to allow a systems administrator to manage files on a computer (at a time when there were generally only a few hundred files). Additionally, when these were developed, computers were not used by the general public, but by highly trained technicians with a thorough understanding of computer technology. The basic paradigm of the tool has not changed in the decades since its introduction, although the user interface to it significantly improved with the widespread introduction of graphical user interfaces in the Macintosh and Windows operating systems. Despite these improvements, the user interfaces of these systems were not designed for modern document management tasks.

A basic principle of user interface design is that the design of a tool should be thoroughly grounded in an understanding of how the users work, what tasks they perform and how those tasks are carried out. However, with personal document management, very little research has been done into how people are managing their documents and what the requirements are for document management tools. This knowledge gap needs to be addressed before better tools can be developed.

2. BACKGROUND

The seminal work in the field of personal information management is Tom Malone's 1983 study titled 'How Do People Organize Their Desks?' [12]. He studied how people used paper files in their offices and identified two distinct strategies: 'neat' and 'messy'. In a neat office, the person tried to designate a category for every document and place it the location corresponding to that category. The location may have been a folder inside a filing cabinet, a paper tray, or a named pile. In the messy office, the person would tend to pile up documents over time, in a less structured way. In both offices, files and piles are the basic building blocks of paper document management.

Several studies have attempted to classify styles of email use in a similar way to Malone's 'neat' and 'messy' classifications. One of the earliest was Mackay [11], who identified 'prioritizers', 'archivers' and 'requesters and responders'. The requesters and responders use email for task delegation; prioritizers concentrate on managing incoming messages while archivers use email to archive information for future use. Whittaker and Sidner [16] also looked at organizing behavior in email, identifying 'no filers', 'frequent filers' and 'spring cleaners'. The 'no filers' were the email equivalent of pilers, allowing all their email to pile up in the inbox, while the filers attempted to place all their emails into folders. The spring cleaners occupied a middle position between the other two groups, using a 'no-filing' strategy most of the time, but periodically attempting to put their documents into files. Without the folders that others use to aid retrieval, 'no filers' rely on full text search and temporal ordering to retrieve their

information. This categorization was extended by Bälter [2] to subdivide 'no filers' in to 'folderless cleaners' and 'folderless spring-cleaners' depending on how often they deleted information from their inbox. A more recent study of email behavior identified two major approaches: 'cleaners' and 'keepers' [10]. Cleaners have specific times for dealing with email, and don't keep events or to-do items in their email. Keepers read email constantly, allowing tasks to be interrupted by email. They keep events and to-do items, and search their email archives.

Studies of organizing approaches taken with respect to web bookmarks have found similar results to the studies of email, identifying 'no-filer', 'creation-time filer', 'end-of-session filer' and 'sporadic filer', depending on whether and when the user saved web bookmarks during a browsing session [1].

The only other more recent study to look at digital documents was recently conducted by Richard Boardman [3]. He analyzed information behavior across three collections: documents, email and web bookmarks with the intention of analyzing difficulties people had in managing their information collections across tools. He found that people could be categorized as either 'pro-organizing' or 'organizing neutral', but that people didn't always adopt the same strategy across all collections. People were more likely to be 'pro-organizing' in their document collection and email than they were in their web bookmarks.

Table 1: Classifications of organizing strategies

Reference	Information Type	Classifications
Malone [12]	paper documents	neat, messy
Mackay [11]	Email	prioritizers, archivers, requesters and responders
Whittaker & Sidner [16]	Email	no-filers, frequent-filers, spring-cleaners
Bälter [2]	Email	folderless cleaners, folderless spring-cleaners, cleaners, spring-cleaners
Gwizdka [10]	Email	cleaners, keepers
Abrams, Baecker & Chignell [1]	web bookmarks	no-filer, creation-time filer, end-of-session filer, sporadic filer
Boardman & Sasse [4]	documents, email and web bookmarks	pro-organizing, organizing neutral

3. RESEARCH DESIGN

This study into personal document management practices consisted of field studies and a survey. In the field study, 10 knowledge workers were interviewed about their personal document management practices. These interviews took place in the participant's offices and participants were asked to demonstrate their structures and processes during the interview. In addition, a snapshot of each participant's file system was taken so that their document structures could be quantitatively analyzed. These interviews were analyzed using thematic analysis and the

resulting conceptual model was used to develop a questionnaire. The questionnaire was used in a survey of knowledge workers designed to gather more generalized data about personal document management practices and to evaluate the conceptual model and personas. The survey was completed by 115 participants, of whom 72 also provided a file system snapshot.

Thematic analysis of the field study interview transcripts revealed three strategies that the participants adopted in order to manage their document: piling, filing and structuring. The three strategies differed in the following attributes:

- Overall level of organization (self-assessment)
- When folders are created (self-reported)
- Preferred retrieval strategy (self-reported)
- Preferred document view (self-reported)
- Use of tree (self-reported)
- Depth of structure (from snapshot)
- Breadth of structure (from snapshot)
- Unfiled documents in top level (from snapshot)
- Folders in top level (from snapshot)

In order to see if these strategies appear in a wider population, a K-means cluster analysis was performed to see if particular combinations of these attributes tended to group together. This analysis was performed using the data from the 72 survey participants who also completed the file system snapshot, and resulted in three distinct clusters. Analysis of variance indicated that several metrics were not contributing to discrimination between any clusters. These included the question on when folders are created, retrieval strategy for old files, use of tree and the breadth of the structure. These were removed one at a time and the cluster analysis repeated until all remaining variables differed significantly across the clusters. Table 2 below shows the resulting variables and the typical values for each cluster.

Using this information and the qualitative information gained in the interviews, user personas were developed for each strategy, and then specific user interface guidelines have been suggested for each persona.

Personas, as defined by Cooper, are "composite archetypes based on behavioral data gathered from many actual users through ethnographic interviews" [5, 6]. Personas provide many benefits, including providing a model of user needs, allowing differentiation between different types of users, and facilitating prioritization of users. Personas help designers with the following tasks: [6]:

- Determining what a product should do and how it should behave. Persona goals and tasks provide the basis for the design effort.

- Communicating with stake holders, developers and other designers. Personas provide a common language for discussing design decisions, and also help keep the design centered on users at every step in the process.

- Building consensus and commitment to the design. With a common language comes a common understanding. Personas reduce the need for elaborate diagrammatic models because, as the authors have found, it is easier to understand the many nuances of user behavior though the narrative structures that personas employ.

- Measuring the design's effectiveness.

They also help prevent the problem of trying to design for all possible users (the elastic user) simultaneously, prevent the designer designing only for him/herself, and focus on the most important interactions, rather than edge cases [6]. For personas to be useful, they need to be grounded in thorough research about the user population.

Some features of personas are:

- Personas are represented as specific individuals.

- Personas represent a class of users in context, not a particular user. The persona encapsulates a unique set of usage patterns.

- Personas have motivations and goals. These provide the fundamental information on which the design is built.

Following these guidelines, the following sections present three personas for personal document management.

4. RESULTS

From the field study and survey data, the three distinct clusters of strategies have been named piling, filing and structuring. The piler strategy identified here is analogous to messy, no-filers, keepers, and organizing neutral strategies identified by other researchers. Filer and structurer are variants of the pro-organizing, frequent-filer and keeper categories identified by others but have some distinct features that mean they are likely to require different user interfaces for optimal support.

Table 2 summarises the results of the cluster analysis:

Table 2: Summary of quantitative features of personas

Metric	Piling	Filing	Structuring
Self reported level of organization	Not very organized	Somewhat organized	Somewhat organized / very organized
Use of search	Last resort	Second choice	Second choice (sometimes first)
Preferred view	List/Details	List/Details	Details/List
Number of Top Level Folders	Medium	High	Low
Number of Top Level Files	High	High	Low
Average depth	Low	Medium	Medium/High

The following sections briefly summarize the main characteristics of each of these three strategies, combining the quantitative data from the survey and the qualitative data from the field studies.

4.1.1 Piling

The piling cluster perceive themselves as relatively disorganized, preferring a list view, with a medium number of top level folders and a high number of top level files and relatively shallow system.

A person adopting a piling strategy doesn't really file his documents; he just lets them pile up in various convenient locations. Folders are usually created in order to dump a large group of old documents that are no longer needed. Because folders are rarely created, the folder structure tends to be fairly shallow, with many folders and files at the top level of the structure. Because recently used files are always easily available, they are retrieved through browsing, with sorting often used to locate the most recent document. A piler may make periodic half-hearted attempts to delete things or organize them into folders, but more because he feels this is how he is supposed to do it than any perceived usefulness. It's peer pressure. Someone adopting a piling strategy tends to be a high Desktop user, since one of the key concerns is least effort and maximum availability. Minimizing visual clutter isn't really an issue, nor does he feel any need or desire to organize documents in order to get an overview of his stuff.

4.1.2 Filing

The second cluster is perceived as more organized, with just in time folder creation, combination of browsing and searching only as a last resort. The structure is medium in depth and width and has a moderate number of unclassified top level folders.

Someone adopting a filing strategy creates folders in order to split up collections of documents. They split folders up if the number of documents grows so large that they cannot easily spot items within them anymore. They tend to create folders either during cleanups or just-in-time as they need to save a folder that doesn't fit an existing category. They do have a hierarchy, although it is moderately broad and not particularly deep. They are likely to have some files in the top level (pending cleanups), and quite a few folders as well, resulting in a tree of moderate depth but high breadth. There is no particular preference for view, but they are much more likely to locate files by browsing their structures than searching. They would generally consider themselves to be relatively organized.

4.1.3 Structuring

Members of the third cluster have high depth, low level of unclassified files, in advance or just in time creation and consider themselves to be fairly organized.

Someone adopting a structuring strategy intensively organizes their files, creating deep and meaningful document structures, often before there are documents to put in them. Related folders are grouped together into more levels of nesting, in order to hide complexity and indicate their relationship. This results in a fairly narrow and deep tree, often with fewer than 3 or 4 top level folders and very few or no files at the top level of their folder structures. They are more likely to browse through their structures although because there are so many folders to inspect, if they can't remember where something is they will readily search, particularly for older files. Browsing is often done using the tree, since the tree gives them an overview of how everything fits together. The parent folders give context to the subfolders. They get frustrated with views that don't show them the full context. For instance, search that only shows them the file name is very irritating. Showing the parent folder is even better, but they really would prefer to see the full path for context. Folders are often created in advance, as soon as a new responsibility, project, course or something appeared on their horizon, to have a place to store the documents. They tend to consider themselves very well organized.

5. PERSONAS

The following three sections describe the three personas created to exemplify these three strategies.

5.1 Piling Strategy (Nathan)

Nathan works hard and plays hard and is always in a hurry. There are never enough hours in the day but he always has some Red Bull handy to keep him going. He drives a fairly old beaten up V8 Holden Commodore. His friends complain that it's always full of papers and junk (especially empty Red Bull cans), but it is certainly fast.

When he gets into the zone, he can be completely absorbed in a task for hours – all he needs is enough Red Bull and maybe some energy chocolate. It's not uncommon for him to look up from his work and discover its 9pm already, and he's frequently late for social events because he got caught up in doing something else. He has one of the messiest desks in the office, since he just doesn't see any value in spending time and effort to file everything properly. His flat is even worse; he can barely see the floor. But as long as he can find clean clothes and anything he's looking for, he sees no real need to tidy up, especially now that he's moved out of home and doesn't have Mum nagging him to do it. After all, any time spent cleaning up is time that could be spent working or playing.

In his office he has piles of paper and books stacked everywhere, but he knows that he can always find anything he needs by going through the pile. If it's something he recently used, it'll always be near the top and it usually won't take him very long to find it. And since his life and his job moves so fast, he doesn't need to go back to old stuff very often anyway.

His computer looks a lot like his office. He usually saves everything on his Desktop because it is one of the easiest places to save things – no thought required. He likes knowing that everything is right there in front of him where he can access it quickly. After all, if he saved it, it's because he needs to do something with it and soon. When he creates files, he usually just lets the application suggest a default filename, since it requires less thought. If he's creating a document, the filename usually ends up as the document title, but sometimes when creating temporary files, he just uses the default filenames such as Document1, Document 2, Book1 and so on. He sometimes wishes he didn't have to bother giving anything names at all. One of his co-workers jokes about him being too lazy to both with proper naming and filing, but Nathan doesn't see it that way – as far as he's concerned, anything that doesn't directly affect the quality of his work isn't considered very important, and that includes filing.

He usually lets things just pile up on the Desktop until he runs out of space. When that happens, he just deals with it the quickest and easiest way he can so he can get back to work. Sometimes he deletes stuff which he is finished with and which he knows has no further use, but usually he just dumps everything except the few most recent active documents into a folder. After all, why spend time deciding which files need deleting and which should be kept? It's not as though he's running out of disk space. Sometimes he wishes that the old stuff would just disappear so he didn't even have to worry about it at all.

When he wants to find a file, he just grabs it from the Desktop. His most recent files are always on the end of the list so it'll be pretty easy to find. If he did a clean up recently, he might have to look in the latest dump folder, but usually what he needs will be on the Desktop. If something doesn't jump out at him immediately or he knows he's looking for an old file, he doesn't waste time browsing around looking for it, but jumps straight into his search tool and enters keywords from the document title. After all, if he's looking for a document, he knows what it is, and he knows that the file probably had a long descriptive title. The faster the search, the better, since all he wants is to find the document and get back to work as quickly as possible. The less thought that has to go into the process, the better.

5.2 Filing Strategy (Linda)

Linda is a very reliable person. Her friends and colleagues know that if you ask her to do something you can safely forget about it, because she will always do anything she promises to. Every morning she gets up at the same time, makes her children's lunches and drives them to school in her old reliable Toyota Corolla. Each night, she makes sure that kids always do their homework and she's never forgotten to attend a parent-teacher meeting or to return a consent slip to her children's schools. Her house isn't super tidy – it's hard with three boys! There is always some clutter around that she hasn't gotten around to cleaning up yet, but it's always clean and has a very comfortable homey feeling. She's fairly methodical and neat, and pretty well organized - she takes a list whenever she goes shopping and she always has her Christmas shopping finished by the first week of December. At work meetings, she's always the one taking minutes, since she can be relied upon to take good notes and to remember to bring them to the following meeting. One of the secrets to her remembering everything so well is her notebook in which she writes all her tasks. She also helps herself out by trying to place things she needs to do something with in areas where she will easily see them and remember.

She tends to be a bit of a hoarder, keeping all her children's baby books and school report cards, as well as their artwork from kindergarten and school. At works she keeps books, magazines and documents from previous years, since she never knows when something might come in useful later on. She customizes her workspace to suit herself, putting the documents she uses most within easy reach, and those she rarely uses on her top shelf, and in the awkward-to-access filing cabinet in the corner under the window. It doesn't bother her if things pile up a bit, but eventually every few months the size of the piles will get too high and she'll have a burst of tidying and put everything away where it belongs. Once something is filed, she doesn't look at it again unless she needs to find it for some reason. She doesn't usually reorganize or clean up material that has already been filed.

She tries to do pretty much the same thing on her computer, tending to have folders for major projects, topics or responsibilities, with all the files related to that task in the folder. Sometimes if a folder gets too big so she has to scroll a lot, she might consider splitting it. While she likes to put things in their correct folder straight away, in practice she often doesn't, saving it in a temporary location first. When she gets a few too many documents in the temporary location she'll go through and file everything properly. Sometimes at the end of a project she might get rid of early drafts or unnecessary files, but she's unlikely to revisit the folder again for cleaning purposes. Generally items enter her folders on a one-way trip.

Linda doesn't like it when the list of folders gets so long she has to scroll. It's like when her physical filing cabinet is packed full

and she can't get anything more in there. She'll take some folders out and put them in the other filing cabinet in the corner, or into boxes on the high shelf. On her computer, she'll either move the folder into an archive folder, or she'll burn it onto a CD and remove it from her hard drive. That way, she doesn't have to see the folder anymore, but the information is still available in case she ever needs it or anyone ever asks her for it.

She usually sees her files and folders through her applications Open/Save file dialog views, and very rarely searches for files. Because all her file and folder names are just such common sense, she doesn't have to look very far to find things. If something was many years ago, she might have to search through her archives, but that would be relatively uncommon. Usually she can find something in a minute or two.

While she prides herself on being able to use her computer pretty well, she doesn't really like it when things change. She found it quite annoying when one of her colleagues changed her file view from the default list to details and then sorted by date while looking at something on her computer – suddenly everything had moved and wasn't in its usual place anymore. She made him change it back. She doesn't like it when things change on her computer without her explicitly taking actions to cause it, being slightly distrustful of things that happen automatically without her knowledge. But she's not a Luddite, she knows that with technology things change, and she's open to making improvements as long as she can see a clear benefit and things don't change too rapidly.

She doesn't think her folder system is anything special. She just splits things up into related groups to make it easier to find things. She's pretty sure someone else could drive her file system if she was away, because she does things that just make sense. She gets into habits of accessing things in certain ways and just uses them over and over again. She isn't really that concerned about using the latest interfaces or cool tools – she doesn't mind if what she's doing is a bit slow or not optimal. As long as it reliably works, it's fine. After all, she drives a Toyota..

5.3 Structuring Strategy (Matthew)

Matthew likes the good things in life – good food, good friends and a good scotch whiskey. It's not about expensive or showy (he can't stand 'bling') but about quality and precision. It's one of the reasons he loves his BMW. He loves the precision German engineering, and the quality and detail that are built into every part of the car. His apartment is very minimalist, with European styling and lots and lots of cupboards. He has a place for everything and everything in its place, and all the places are hidden behind cupboards and frosted glass or concealed in staircases. All his kitchen appliances are hidden away in appliance cubbies. The only thing visible is his deluxe coffee machine both because he uses it so frequently and because he deliberately bought one with the least cluttered and smoothest exterior. Although he tries to keep everything clean and tidy all the time, in reality things can get a little bit untidy for a few days before he tidies up again.

At work he tends to have a pretty clean desk. He has several organizer boxes on his shelves and paper trays on his desk. He prefers to always have everything in its place so he knows exactly where to find it. He has a weakness for organizing systems – he just can't go past Howard's Storage World without buying something. He recently bought a deluxe labeling machine so he can put labels on all his shelves and boxes, and got a range of

different colored labels to go with it so he can use colour codes. His friends sometimes tease him about being excessively anal about filing.

Matthew likes to be just as organized on his computer. His Desktop is clean – just shortcuts to the one or two items that he accesses all the time, which he changes depending on what he is currently working on. He tries to give every file a descriptive name, sometimes with codes to indicate the year, project or task, and placing it in its proper place in the hierarchy.

Whenever he starts a new project, one of the first things he does is select a name for it and create a folder structure (often similar to previous projects). If he knows in advance what documents he'll need to create, he'll often create the outlines for those in advance, making sure they all use the same template so all the styles and formatting are consistent. If he knows he'll be corresponding with others about it, Matthew will make a folder in his email system for it, and if he knows he'll be doing a lot of web searches, he'll make a folder in his browser's favorites to store all the related web links. If he's going to be working with paper documents or books, he'll create a label on his shelves to contain the related material.

Matthew sometimes switches between different ways of organizing things – it's important to be organizing things as well as he can and he's never sure if he's doing things the best way. He wonders, should he keep his trip expense reports in a separate expense report folder or should he put them in the folder with the rest of the information about the trip? He wishes he could have things in more than one place. And with some of his projects now spanning multiple years, he's never sure whether he should create year folders inside project folders or the other way around. He currently has year folders as the top level, because he stumbled across an article on the web which argued that this was more efficient.

He likes the fact that the tree view gives him an overview of the structure of his project but he wishes it was more useful, like letting him know which parts he still needed to work on and which were completed. He downloaded a trial shareware application that let him colour code his folders, which he liked for a while, but didn't like quite enough to pay $50 for it. He's also downloaded a couple of shareware applications that present different views of his folder structures, showing how the parts of the structure are related to each other and letting him follow links from one folder to another.

When someone else sees Matthew's folder structures he feels that they aren't just seeing the places where he stores his files, they're seeing the structure of his mind. He just wishes he could make his folders a little bit more expressive of his own mental representations.

6. USER INTERFACE GUIDELINES

The following table lists some specific user interface guidelines based on the unique characteristics of each persona. The following sections describe these in more detail.

Table 3: Strategy-specific user interface guidelines

Piling strategy (Nathan)
Do not require containment
Support a time based interface

Provide optional tagging
Filing strategy (Linda)
Support containment
Provide a cleanup interface
Support reminders
Structuring strategy (Matthew)
Support containment with multiple classification/dynamic containers
Provide optional relationships between items
Provide optional tagging and colour coding
Provide optional custom metadata

6.1 Piling Strategy (Nathan)

6.1.1 Do not require containment

Nathan doesn't need a folder-like containment mechanism in order to group his documents, since he is interested in expending as little up-front effort as possible. This doesn't mean that folder or a grouping mechanism needs to be completely absent, just that if present, it should be optional. It should be entirely possible to use the interface without ever having to think about where to put something or in what to contain it. The 'out of sight, out of mind' principle isn't one that Nathan subscribes to.

The attempt to take literally the piling paradigm to create a user interface that supports piles is misguided when it comes to the personal document management piler. Electronic implementations of piles (e.g. [13]) are a containment mechanism just like folders. Conceptually, they operate exactly as folders although with a slightly richer visual representation, one which folders views could easily match (and with picture folders starting to show thumbnails of contents, this is getting closer). The nature of the piling strategy is that followers don't really want to group and organize things. He adopts piling because it involves the least initial effort.

This doesn't mean that a containment or grouping or folder needs to be completely absent, just that if present, it should be optional. It should be entirely possible to use the interface without ever having to think about where to put something or what to contain it in.

6.1.2 Support a time based interface

Time based retrieval is more important to users of a piling strategy than users of other document management strategies. The piler naturally has (or maybe is forced to have) some sense of chronology, since their pile stacks up in order of creation/acquisition. While they don't need to remember absolute times or time spans, they need to have a relative idea how far back through the Desktop stack to look, or how many cleanup folders back to look for something. An interface such as Lifestreams [9], provided it had very strong search support, would probably suit the piling strategy very well.

One way of leveraging this tendency is to ensure the default document view shows all recent files ordered by either when they were most recently used or when they were created. The Desktop could potentially use the same view, making the view easier to access. This view should be dynamic, rather than the static view currently offered by the Desktop. Items that have not been used recently should just disappear from view. Thus, the default view might show an item that was added a month ago but which was used three days ago, while an item added two weeks ago but not used since may not be visible.

Rather than having items disappear after a certain time, the view should simply show as many recent documents as possible. This takes advantage of the common practice of sorting by date to find the most recent document, and eliminates the need for periodic cleanups or dumps of files. There should be an option to 'jump back' or scroll back to show earlier sets of documents as well, giving this interface something in common with the TimeScape software [14], although without the spatial element.

All dates and times should be shown as relative times by default (although the option of switching to absolute times should be available), since few people have sufficiently good recall to pinpoint exactly when they created or worked with a document. Examples of relative times include '30 minutes ago', '5 hours ago,' 'yesterday' and '2 weeks ago.'

6.1.3 Provide optional tagging

If someone adopting a piling strategy wants to do any kind of categorization at all in order to make sure that he is more easily able to retrieve stuff, the easiest way to support this would be to allow tags to be specified when saving the document (or added later). These can be free-form comma separated tags in which he can just type additional keywords that he might want to use to search for it but that don't appear in the document itself. This provides a way of being able to group related documents without the containment semantics, since it is easy to create a view of all documents sharing the same tag or tags. The advantage of tagging is that it lets the user add words they associate with the documents, but which might not appear within it. This makes future searching more effective.

6.2 Filing Strategy (Linda)

6.2.1 Support containment

Users adopting a filing strategy need a containment mechanism in order to group their files into manageable locations. The standard folder metaphor would probably work very well, although there are many other ways of implementing containment semantics, which would also work. Different means of visualizing containers should be explored. One place to start would be developing views that allow more of the hierarchy to be seen at once, since a common complaint is the time taken to click down the levels. Within containers, items should be able to be viewed with or without details, since name is probably the most important dimension. If other dimensions are used, file type, date created and date last used would be the most useful.

While the ability to change sorting is important, there should also be a custom sort or user defined sort. In this way, filers could organize things into the exact order they wanted and know things wouldn't change. This creates a sense of stability and permanence and makes finding items through known paths easier and more reliable. It also obviates the need to change the 'common sense' file names in order to force a sort order.

6.2.2 Provide a cleanup interface

While most of the time a user of this strategy is interacting with documents through Open/Save dialog boxes, they do want a larger view of their file structure when doing a cleanup. During a cleanup, they are going through files in a temporary location (e.g. top level folder or Desktop) and placing them into their permanent folder home. To do this effectively they need to be able to see the list of files they are cleaning up, as much of their folder structure as possible (expanded tree view), and ideally a preview, in case they need to be reminded what the document is before they can decide where to put it. In this view it must be easy to create new folders and to reorder folder contents in the tree.

6.2.3 Support reminders

Being able to place documents somewhere she will be easily reminded of them would be a good feature for this strategy. A useful addition would be the ability to explicitly set a reminder on a file, which could then be used to pop up a reminder message at a certain date and time. It could also alter the appearance of files that had reminders attached so they were more visually obvious.

6.3 Structuring Strategy

Followers of a structuring strategy need the ability to express containment just as filers do, but they also need richer containment semantics.

6.3.1 Support hierarchical containment with multiple classifications

Systems must provide the ability to create hierarchies of containment, since many people appreciate the ability to create folder structures. Multiple classifications enable a document to live in more than one location. Previous means of approximating this such as shortcuts or copies are not sufficient – the document actually needs to have one location but appear in multiple locations. Regardless of the location from which the file is viewed and accessed, any changes to the document or its metadata should be immediately effective in all locations. When a file is deleted, if it exists in multiple locations the user will need to be prompted whether the file should be deleted from that location only or from all locations.

A user interface should support collapsing or hiding of levels of information, to enable the ability to see an overview and drill down to detail on demand.

6.3.2 Support dynamic containers

Providing dynamic containers is another way of providing some of the same functionality as multiple classifications. Dynamic containers don't have a predefined set of contents, but rather display the contents based on a search. The containers in the Presto system [7] are an example of this, as are Outlook 2003's Search Folders. For instance, Matthew's expense reports could be stored in the folder with the rest of his trip information, but he could create a dynamic folder that presents a view of all his expense reports together. The dynamic folder can be organized into folders like any other folder.

6.3.3 Provide relationships between items

To a structurer, the file system is more than simply a place to store things; it is a representation of the structure of his information. For this reason, the ability to make arbitrary relationships between things would be a useful extension. This can be partly automatic and partly manual. For instance, the system could track which documents are opened with other documents or emailed together with other documents and therefore infer relationships between documents. This could be presented by having a 'Related items' panel that displayed the other documents related to the currently selected document, enabling them to be quickly accessed. In addition, there should be an ability to manually create relationships between items, thereby choosing the items that appear in the 'related items' view.

6.3.4 Provide optional tagging and colour coding

Other methods to provide the structuring filer with richer abilities to organize files include allowing the ability to tag documents or files with keywords (as described for Nathan), and to colour code files and folders. These should be entirely optional but if used are entirely user-generated. The organizer can use any colours they want, and can assign an optional descriptive label to the colour, or just simply use the colour.

6.3.5 Provide optional custom metadata

The 'Rolls Royce' of systems for an organizer would be to allow them completely free rein to construct their own properties to be added to files and folders and to use these properties to create dynamic folders and hierarchies. These could then provide the basis of a customized search function that provides a means of finding information by arbitrary metadata. Whist this provides the ultimate in flexibility, it requires considerable effort and overhead to maintain, and it must be acknowledged that relatively few users are interested in organizing to quite this extent.

7. DISCUSSION

Some results from the classification model differed from the strategies described based on the field study. For example, it was anticipated users of a piling strategy would make greater use of search tools to compensate for their lack of folder structure. However, it is possible that their piling strategy means that most of the time they can browse through their top level documents, assisted by sort options until they find their target document. In this way, they are predominantly relying on a browsing technique rather than search. In contrast, adopters of a structuring strategy were not expected to be heavy users of search, since the effort they expended in structuring their folders should pay off by providing more effective browsing. However the survey results showed that structurers were more likely to search in their own documents. This result has subsequently been independently observed in a study of email [15].

It is unclear whether more frequent searches mean the document management system is less effective. It is possible that the folder hierarchy makes the search much more useful through being able to search only a related subset of the documents, and because the metadata provided by the folder path makes recognizing found documents easier. More research would need to be done examining the amount of time spent in document management activities by adopters of the various strategies before a determination can be made.

It was also anticipated that adopters of a piling strategy would be much less inclined to use the tree, but since the question about tree use didn't ask for frequency, there is no way of knowing whether they use it as much as users of the other strategies.

8. CONCLUSION

This paper has described the development of a model of three document management strategies: piler, filer and structurer. A rich user persona was created for each in order to exemplify typical behavior. This persona can be used as a tool to aid the design of new interfaces for personal document management. In addition, some specific guidelines were presented for user interfaces to support each strategy.

It is necessary to remember that although these strategies and the personas that illustrate them are useful tools to guide user interface development, people do not necessarily neatly fit these three strategies all the time. People will at times adopt one or the other depending on the circumstances, although there is usually a dominant preference. These three categories collectively cover the spectrum of personal document management behavior observed in this study and therefore an interface that can accommodate all three should be useful to everyone.

9. REFERENCES

[1]   Abrams, D., Baecker, R. and Chignell, M., Information Archiving with Bookmarks: Personal Web Space Construction and Organization. in *CHI'98 Conference on Human Factors in Computing Systems*, (Los Angeles, California, USA, 1998), ACM Press.

[2]   Bälter, O. Strategies for Organising Email *Proceedings of HCI on People and Computers XII*, Springer-Verlag, 1997.

[3]   Boardman, R. Improving Tool Support for Personal Information Management *Dept. of Electronic and Electrical Engineering*, Imperial College, London, England, 2004.

[4]   Boardman, R. and Sasse, M.A., "Stuff Goes into the Computer and Doesn't Come Out" A Cross-tool Study of Personal Information Management. in *CHI'2004 Conference on Human Factors in Computing Systems*, (Vienna, Austria, 2004), 583-590.

[5]   Cooper, A. *The Inmates are Running the Asylum*. Sams, Indianapolis, Indiana, USA, 1999.

[6]   Cooper, A. and Reimann, R. *About Face 2.0*. Wiley Publishing, Indianapolis, USA, 2003.

[7]   Dourish, P., Edwards, W.K., LaMarca, A. and Salisbury, M. Presto: An Experimental Architecture for Fluid Interactive Document Spaces. *ACM Transactions on Computer-Human Interaction*, *6* (2). 133-161.

[8]   Faichney, J. and Gonzalez, R. Goldleaf Hierarchical Document Browser. *Australian Computer Science Communications*, *23* (5). 13-20.

[9]   Fertig, S., Freeman, E. and Gelernter, D. "Finding and Reminding" Reconsidered. *SIGCHI Bulletin*, *28* (1). 66-69.

[10]   Gwizdka, J., Email Task Management Styles. in *CHI'2004 Conference on Human Factors in Computing Systems Extended Abstracts*, (Vienna, Austria, 2004), 1235-1238.

[11]   Mackay, W.E., More than just a communication system: diversity in the use of electronic mail. in *CSCW'88 Conference on Computer-Supported Cooperative Work*, (Portland, Oregon, USA, 1988), ACM Press, 344-353.

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

[13]   Mander, R., Salomon, G. and Wong, Y.Y., A 'Pile' Metaphor for Supporting Casual Organization of Information. in *CHI'92 Conference on Human Factors in Computing Systems*, (Monterey, California, USA, 1992), 627-634.

[14]   Rekimoto, J., Time Machine Computing: A time-centric approach for the information environment. in *UIST'99 Symposium on User Interface Software and Technology*, (Asheville, North Carolina, USA, 1999), 45-54.

[15]   Teevan, J., Alvarado, C., Ackerman, M.S. and Karger, D.R., The Perfect Search Engine Is Not Enough: A Study of Orienteering Behaviour in Directed Search. in *CHI'2004 Conference on Human Factors in Computing Systems*, (Vienna, Austria, 2004), 415-422.

[16]   Whittaker, S. and Sidner, C., Email Overload: exploring personal information management of email. in *CHI'96 Conference on Human Factors in Computing Systems*, (Vancouver, Canada, 1996), 276-283.