

Filing, Piling & Structuring: Strategies for Personal Document Management

Sarah Henderson
University of Auckland
s.henderson@auckland.ac.nz

Ananth Srinivasan
University of Auckland
a.srinivasan@auckland.ac.nz

Abstract

Personal digital document management describes the activities performed by an individual in creating, acquiring, organizing and maintaining collections of their documents. The abundance of digital information that knowledge workers today must manage means that providing useful and usable tools to organize and handle this complexity is more important than ever. In order to better understand the differing approaches people take to document management, a study was conducted involving a combination of in-depth interviews and a survey. Qualitative analysis of the data from the interviews and quantitative analysis of the data from 72 survey participants 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. Understanding of these three strategies can be used to inform the development of better tools to support document management.

1. Introduction

There are many individuals for whom information processing (especially digital information) is now a significant part of their jobs. Peter Drucker coined the term 'knowledge workers' in 1959 to describe this group of people [9]. For these people, information is no longer a scarce resource; information exists in abundance and human time and attention has now become the scarce resource [24]. Information overload is now a recognized problem as people struggle to manage the increasing quantities of information they need to deal with on a daily basis [11].

Even as these knowledge workers use the growing range of data processing tools and visualisation software to manage the avalanche of data, there is another overload creeping up on them. A morass of reports, memos, articles, notes, presentations, graphics, contacts, web URLs, emails, tasks and appointments has slowly but surely been accumulating on their computer. While finding information in databases and

on the web is becoming easier, finding information located on their own hard drive is becoming more and more difficult.

Most people store their information in the hierarchical file system provided by their computer's operating system, and manage their documents through a hierarchical file browser such as Windows Explorer [12]. These file browsers were originally 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 such as those commonly used in the Macintosh and Windows operating systems. Despite these improvements, the user interfaces of these systems were not designed for modern document management tasks.

Managing documents is a challenge for organizations, groups and individuals. At the individual level, personal document management is term used for the set of activities involved in managing a collection of digital documents. 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. Because it is so pervasive, small improvements in document management efficiency and effectiveness can scale up to large overall productivity gains.

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 [17] study titled 'How Do People Organize Their Desks?' 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 [16], 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 [27] 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 [3] to subdivide 'no filers' into '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'[14]. 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 [6]. 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 summarizes the different organizing strategies observed in previous empirical research in this area.

<i>Study</i>	<i>Information Type</i>	<i>Classifications</i>
[17]	Paper documents	neat, messy
[16]	Email	prioritizers, archivers, requesters and responders
[27]	Email	no-filers, frequent-filers, spring-cleaners
[3]	Email	folderless cleaners, folderless spring-cleaners, cleaners, spring-cleaners
[14]	Email	cleaners, keepers
[1]	web bookmarks	no-filer, creation-time filer, end-of-session filer, sporadic filer
[7]	documents, email and web bookmarks	pro-organizing, organizing neutral

3. Research Questions and Method

The fundamental issue that we address in this work is the examination of document management strategies that individuals employ in the context of their work. Specifically we ask a number of questions. What are the patterns of document management that individuals exhibit? Are there determinants of why there are variations in these patterns? How can the discovery of these patterns assist in improving the usability of tools? In order to answer these questions, we undertook a study to investigate the patterns of usage among workers in a university environment. Such an environment is particularly helpful for work of this

nature because it encompasses a wide variety of usage situations coupled with a good mix of individuals with varying requirements.

The study utilized two complementary approaches to data collection. The first is a series of in-depth interviews of ten individuals. The results from analyzing the data from this phase led to the design and administration of a survey to a larger set of users.

In the first part of the study, ten knowledge workers were interviewed about their personal document management practices. The interviews were semi-structured, and were carried out in their own offices so that their document management practices could be viewed in their natural context. The participants were encouraged to demonstrate their document structures and work practices to the researcher rather than simply explaining them. The interviews were guided by an initial set of questions raised by previous research, and this guide sheet was updated between each interview to include new topics raised by each participant.

The semi-structured nature of the interviews meant that the concerns and processes of the participants were able to emerge. This technique of interviewing participants in their offices and using their computers as a questioning point for the interview has been used many times in investigation of related aspects of personal information management [10,17,26-27], and was used in the only prior studies of personal document management [4-5].

Although contextual information from the environment was included, the main source of data was the self-reported perceptions and opinions of the person involved. This is very good for understanding how they view the issue and what problems they encounter, but cannot provide information about exactly what they do, and what structures they create. There are a number of reasons why self-reported experiences are not always accurate, including imperfect memory [21], wanting to give an answer that is socially acceptable [23] and making different assumptions to the researcher about what is important [22]. These issues have also been found in usability studies, where what people say they do does not always exactly match what occurs [18]. For this reason, objective information about their document structures was also collected through use of a file system snapshot tool.

The context of personal document management is a person and their computer, and therefore this is the natural unit of analysis. It was decided to focus on the work setting because it is in a work context that the issue of information overload has been raised, and it is in a work context that the issue of productivity declines as a result of ineffective information retrieval have been identified [13]. It is expected that the work

situation is the highest use, and has the most negative consequences if document management practices are inefficient. The participants selected were all users of the Microsoft Windows operating system. This was partly for practical reasons, since different operating structures store files in different ways, and obtaining tools to take a snapshot from any operating system is likely to be extremely difficult. The decision was justified by the fact that at the time of the data collection, the Windows operating system has approximately 93% of the desktop operating system market [19], and is more prevalent in organizations than in home settings [15]. The selected sampling frame is the staff members of a University faculty. Table 2 summarizes the interview participants.

Table 2. Interview participant summary

Alias	Role	Gender	Age Range
Alex	Technical staff	Male	20-29
Brett	Course manager	Male	20-29
Candice	Course manager	Female	40-49
Damien	Course manager	Male	20-29
Edward	Junior academic	Male	50-59
Frank	Junior academic	Male	50-59
Gail	Junior academic	Female	Unspecified
Harriet	Mid level academic	Female	40-49
Ina	Senior academic	Female	40-49
Jack	Senior academic	Male	60-69

These interviews were analyzed using thematic analysis, and an initial conceptual model of document management concerns was developed. Thematic analysis is a process of coding qualitative information [8], that focus on identifying themes and patterns of behaviors [2]. The transcript was first examined for any major themes and patterns that emerged. This was followed by an iterative process of examining the transcript for any data that relate to the already identified patterns, while being alert to new patterns that emerged. The patterns were then further classified into subthemes. As much as possible, the theme and subtheme names were taken from the actual words and phrases used by the participants [20].

The interviews were analyzed and coded in the order they occurred. Initial coding was done at a very specific level, with coding for things such as an aversion to search, preferences for time based sort, or reluctance to delete. After the initial coding of each participant using the ideas and themes that naturally emerged, the themes were examined and related ideas grouped together as subthemes of a common theme (such as search, sort, delete behavior). Additionally, the previous interviews were reviewed again to identify any instances of the newer themes. This was repeated through all the interviews until finally arriving on a set of stable themes and subthemes.

4. Interview Results

The 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 six distinct dimensions. These dimensions are described in the following sections.

4.1. Overall level of organization

Alex describes himself as very disorganised, and later remarked *“I fully admit that I’m pretty hopeless at managing files.”* Brett also considers himself pretty disorganised.

Candice rates herself as average. While she doesn’t have any particular difficulties, she does wish she had time to clean up her files, and worries that sometimes she creates folders for documents when there is probably already an appropriate folder somewhere if only she could remember where. Frank also gives himself an average rating, saying *“I try to organise my files, it is not easy, it is easy to say, but difficult to do, how to well organise the file structure on my computer.”* Gail considers herself averagely organised, and although she sometimes feels that she should be more organised, it doesn’t really bother her. Jack gives himself an average rating, saying that he was initially *“very conscious and careful about producing clear directory structures and then latterly I just shove files anywhere,”* although his search tool means that this lack of organisation doesn’t impede him finding documents quickly.

Ina rates herself as above average in organisation, but says her rating varies with time. After a reorganisation, she would rate herself as extremely well organised, but she lets it drop down to a fairly disorganised state and then thinks ‘oh I can’t stand this anymore’ and reorganises. She notes that this is *“my*

perception of what’s organised in terms of a system that works for me.”

Damien rates himself as very well organised, saying *“I’m not anal about it but it’s fairly organised. I know where everything is usually.”* Harriet also rates herself highly, saying, *“Most of the time I actually know what my files are. It’s scary, but I can usually find what I want immediately because I take care in using consistent naming.”*

While this is a very subjective assessment, after taking a tour of the participant’s files and examining their file system snapshots, the researcher believes all the participant’s self-assessments are accurate.

4.2. Preferred retrieval strategy

The majority of the participants reported that if they need to locate a document, they would browse to it in their folder structures. Brett calls this a *“manual search”*, saying *“Normally there’s only going to be a couple of places it’s going to be. One of maybe three places.”* Candice also says she’d go and look in her normal structure, and depending on what kind of file she was looking for, she’d know where to go and look. Damien also says he usually knows the whereabouts of all his files. He says *“I usually know where I put stuff.”* He doesn’t see finding his documents as a big problem except that it takes a little time: *“You know where they are but you still have to click through to get them.”*

Edward says he can browse back to the folder if he remembers it: *“Normally, I just go to ... if I remember clearly where I put there, then I go to for example My Document folder to find it, or other folder to find it.”* Frank makes this process easier on himself by having a standardised organisation structure across all his computers and between his email and document folders. Because he is so familiar with his folders, he says *“if I’m looking for something, the zone of possibilities is quite small, and I can find material relatively efficiently.”*

Gail usually knows where her files are stored because she feels the structure she has created makes sense for her. Harriet finds it easy to find files again: *“Most of the time I actually know what my files are. It’s scary, but I can usually find what I want immediately because I take care in using consistent naming.”* Ina takes pride in being able to locate her documents immediately, both physical and digital documents. She says files are in the first place she looks about nine times out of ten, and says that, *“people say that they are surprised that they can come into my office and they can ask for an article and I will know where it is”*.

Jack browses for a document if he can remember exactly where it is, otherwise he resorts to a search tool. Similarly, Alex will find a document by browsing if it is on his Desktop or if he knows for certain it is in his most recent cleanup folder, otherwise he will search for it.

Alex is a very frequent user of search. On numerous occasions during the interview when he wanted to show some particular aspect of his organising scheme or naming scheme, he would open the Windows XP search tool and search for the files or folders he wanted to show. Almost all his searches are based on locating keywords from the file name.

While Frank prefers browsing if he can remember the location of a document, he very often resorts to searching for it: *"I told you that sometimes a mess, I try to find some file, so I use a lot of search."* His preferred method of searching is to first try keywords from the filename, and if that doesn't work, to move on to searching text within the file: *"I try to remember the keywords. For example, first of all I try to search the documents, the folders, by remember some keywords for the folder name. If unfortunately I could not find that folder or the file name, I use search for the text within the file, to search that. Sometimes it work very well."*

Jack uses Copernic Desktop, a third party search tool in order to locate his files. While he does browse to something if he knows exactly where it is, he frequently turns to Copernic to find his documents. He describes it as a *"Godsend"*, because he doesn't have to adhere to his clear directory structures but can just *"shove files anywhere"* and know that he can *"just pull them back by searching for them without having to know where they are."* He primarily searches by file content, not the file name: *"It's that content which I tend to remember, people maybe with more logical minds or more structured minds will recognise that it's this type of topic and therefore it's in this folder for this topic, but because my folders got rather messed up, and I can't be bothered to get back and sort them out, I now just don't worry too much."*

He also appreciates the preview feature that Copernic offers (which the standard Windows XP search doesn't have), allowing him to easily see if he has found the right file or not.

Brett rarely searches for documents, doing so only if he can't remember where he put a file, often because he's moved it somewhere else on the computer. His searches usually don't narrow down the location, because *"if I'm using search in the first place it's because I don't know where something is, so I'll search the entire hard drive."* He normally searches using keywords from the file name, and if he knows the type of the file, will include the extension as a

keyword. Candice and Gail also say for them search is a last resort if they haven't been able to find something by browsing documents, and they'll also use keywords from the file name.

Frank also rarely searches his documents: *"Occasionally I do, but by having a standardised structure, more often than not the document is where I think it should be."* His use of replication tools also means when he does search, he only needs to search one document collection, rather than wonder which of his multiple drives might contain the document.

Damien almost never has to search for something in his own documents: *"I don't use the search function as such, because I usually know whereabouts they are."* *"But I suppose, rarely, I'd say I use the search, but sometimes, I suppose you might lose something."* If he had to search, it would be using keywords from the file name.

Harriet has a similar perspective, saying the only time she needs to search is occasionally when *"I've saved an attachment without paying attention, yeah, occasionally I'll do that. It saves to the default temp folder and I think, 'I'm sure I saved this in the right place' and so I'll go and search and it's invariably in the temp download folder."*

Ina has never needed to search for anything:

"My husband tells me I should use it but I haven't lost anything badly enough to need to do that. I'm pretty good at finding what I've lost. Really I could count on maybe one hand if I've ever thought 'where the hell is that file?'. I just haven't run into that. And that's like forever, so I don't think I know how to use that Search File properly. Whereas the way my husband files, he's always searching."

4.3. Preferred document view

The two views most commonly used by participants were the list and details views, with icons only used by one person, and disliked by three people.

Edward's preference for the details view is *"because sometimes I need to check the size of the file and also the date. And also sometimes I try to find out or organise my file by using the type, so these three attributes are quite often used by me."* Brett also says he uses the detailed view whenever he can, and Candice says that if she ever encountered a folder in another view, she'd change it to details. Gail's preference is due to the ability to see more information at a glance.

Jack's preference was for details view:

"I much prefer to see a details list of files, I don't like icons." He didn't feel other people necessarily shared this preference however, saying *"most people do like icons and if they were to say because of that"*

they were going to get rid of the details list I would think that was a retrograde step.” Alex also explains that because he so often works with lists, he always uses the details view: “I hate viewing it in terms of icons.”

Harriet also says “I hate icons,” preferring normally to use the list view to see her documents, although “occasionally I’ll look for details, particularly with my online course, I’ll sometimes look at that with details because that has a lot of things I had scanned into PDF form, and in details I can pick up its size.”

Damien had the details view in place for all his documents, but explained that he normally uses the list view. When asked why it was in details view, he replied:

“Because I haven’t bothered changing it. Cause I don’t have a ... OK, well, at home it’s very different, cause at home I have a lot more information. That’s actually, yeah. I think cause I don’t have enough in here that it goes onto multiple screens. As soon as it hits multiple screens then I go to list so I can see everything. But here I probably have it on detail just cause it hasn’t annoyed me. Hmm that’s actually interesting, because I don’t actually use list the most here, I use details, quite right.”

Ina was the only one to use the icons view at all, and uses a combination of list and icons depending on how she is accessing her files. If she is using My Computer, everything is in icons view. However, her primary mode of accessing files is through the Open/Save dialogs in Microsoft Office applications, and there, due to the small size of the window, she changes it to list view so she can scroll through more easily. Her use of My Computer and the icons view is usually reserved for when she is doing a reorganisation.

4.4. Depth of structure

Systems vary in the level of nesting of folders. Shallow structures may only have one or two levels of nesting, whereas deep structures can have many nested folders. It is useful to know the maximum and average depths of the folder tree, not least because this may affect what would be appropriate visualisations of the tree structure. The top level folder in a structure has a depth of 0. A subfolder of this structure has a depth of 1; its subfolders have a depth of 2 and so on. The depth of a participant’s file system is the number of levels deep their folders are nested. There are two ways to look at this – (1) the maximum level of nesting they have anywhere in their system, or (2) the average depth of each folder’s depth. Average depths tend to be between 2 and 6 levels. The deepest folders in a participant’s folder system tend to be between about 5

and 12 levels deep. As you’d expect, there is a close relationship between the average and maximum depth values. The participant’s depth self-assessments were very accurate, with Ina, Harriet, Alex and Jack all stating that they tend to have shallow structures. These four participants had the lowest average depths.

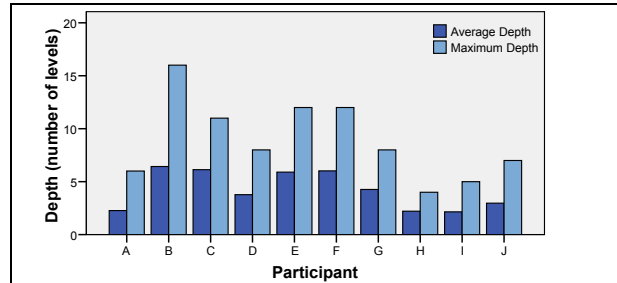


Figure 1. Average and Maximum Depth

4.5. Unfiled documents in top level

Documents are often left unfiled in the top level, for instance, on the Desktop or directly in the My Documents folder. The proportion of files in the top level of the hierarchy provides useful information about this tendency. The proportion of folders at the top level of the tree also indicates how top-heavy the tree is. Harriet and Alex again have the highest proportions in these locations. Harriet’s percentage is exaggerated by the small size of her document collection – while she only has 26 top level files, she also only has 372 files in total. Alex’s smaller proportion reflects 188 files at the top level of his file system.

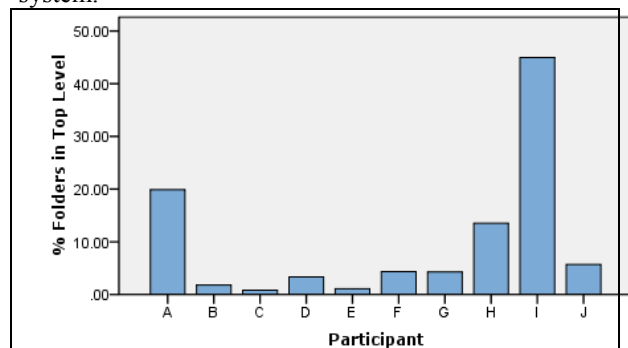


Figure 2. Proportion of Folders in Top Level

4.6. Folders in top level

Folders in the top level are on the Desktop, inside My Documents or inside whatever other folder or device is used as a primary storage location. It is notable here that Ina and Alex have the highest

proportions of folders in these locations. For both, this is a result of their shallow and broad hierarchies.

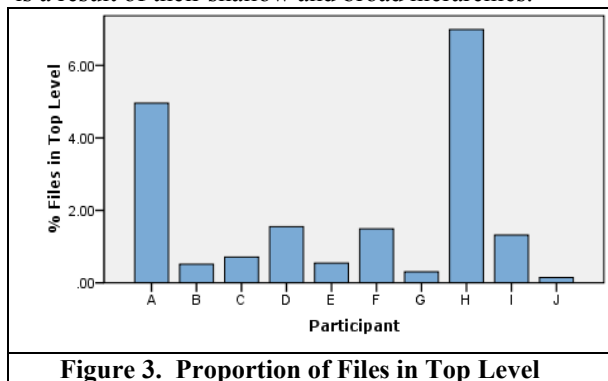


Figure 3. Proportion of Files in Top Level

5. Strategies

From the initial analysis of the interview and file system snapshot data from the field study, the participants seemed to loosely adopt one of three main strategies: piling as practiced by Alex, structuring as practiced Damien, Frank and Harriet, with the remainder adopting a filing strategy as necessary to ensure retrieval.

5.1. Piling

Alex doesn't really file his documents, he just lets them pile up on his Desktop until it is full and he dumps them into a folder. Because of this he has a fairly shallow and broad file system, with a higher proportion of documents at the top level of his structure. He doesn't really use the tree (since he doesn't really have many folders organised into a hierarchy) and is more likely to search for lost files if he can't find them using chronology. He is also likely to use the details view more often so he can sort items easily. He rarely creates items in advance and because of this lack of structure, considers himself relatively disorganised.

5.2. Filing

Most participants file documents into folders in order to retrieve them later. 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 file 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. 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 organised.

5.3. Structuring

Structuring filers such as Damien and Frank share many of the same behaviours as filers, but for them the creation of a structure serves a larger purpose than simply retrieval. It also provides them with an overview of the structure of their information, and for this reason they will often create folders in advance of having files to put in them, simply because seeing the conceptual categories is useful. They are likely to have fairly deep and narrow structures, and to have relatively few documents at the top level of their folder structures. They are more likely to browse through their structures although because there are so many places to look, will more readily search for older files. They tend to consider themselves very well organised.

6. Survey Results

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

In order to see if these three strategies appeared in the wider population, a K-means cluster analysis was performed to see if particular combinations of these dimensions 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 3 summarizes the results of the cluster analysis performed on the survey data.

The clusters identified in the survey data map onto the piling, filing and structuring strategy behavior identified in the interviews. The piler strategy identified here is analogous to messy, no-filers, keepers, and organizing neutral strategies identified by other researchers.

	Cluster 1:Piling	Cluster 2:Filing	Cluster 3:Structuring
Number of Participants	12	29	31
How organised?	67% not very 25% somewhat	76% somewhat 24% very	58% somewhat 36% very
Use of search	67% last resort 25% second choice	65% second choice 31% last resort	61% second choice 26% last resort 13% search first
Preferred view	50% list 42% details	42% list 42% details	77% details 20% list
Number of Top Level Folders	67% medium 33% high	69% high 28% medium	77% low 23% medium
Number of Top Level Files	58% high 42% medium	55% high 41% medium	77% low 23% medium
Average depth	80% low 20% medium	75% medium 25% low	54% medium 45% high

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. There is no difference between the groups in people's overall satisfaction with their document management practices.

7. Discussion

Table 4 and subsequent 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.

6.1. 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.

	<i>Piling</i>	<i>Filing</i>	<i>Structuring</i>
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

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.

6.2. Filing

The second cluster is perceived as more organized, with just-in-time folder creation, browsing as a primary document location tactic 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.

6.3. Structuring

Members of the third cluster tend to have deep structures, a low level of unclassified files, employ 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 if they can't remember where something is they will readily search, particularly for older files (because there are so many folders to inspect). Browsing is often done using the tree, since it 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 appears on their horizon, to have a place to store the documents. They tend to consider themselves very well organized.

6.4. Structure vs Strategy

It is clear from the classification model developed that users of different strategies have different structures. Since workspaces with different structures will provide different levels of support for cognitive offloading, and will have different cost structures to access information within the workspace, the structure of the workspace will necessarily influence the strategy that is adopted. Thus strategy and structure are inherently interrelated. Changes in the strategy will influence the structure and changes in the structure will influence the strategy.

Just as with biological evolution, it is entirely possible that there are multiple stable and successful combinations of strategy and structure, and there may not be a single optimal solution. The solution adopted may be influenced by individual differences, or on chance events such as inheriting another person's filing system. More research is necessary to identify the factors influencing the development of particular strategy/structure combinations.

Some results from the classification model based on the survey data differ from what might be expected based on previous information management classifications. In particular, earlier studies of email [3] suggested that that users of a piling strategy would make greater use of search tools in order to compensate for their lack of folder structure. By contrast, in this study a preference for location based search was observed. Therefore, 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 also been observed in a study of email [25].

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. 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] D. Abrams, R. Baecker and M. Chignell, Information Archiving with Bookmarks: Personal Web Space Construction and Organization, *Proceedings of the CHI'98 Conference on Human Factors in Computing Systems*, Los Angeles, California, USA, 1998.
- [2] J. Aronson, A Pragmatic View of Thematic Analysis, *The Qualitative Report* 2 (1), 1994.
- [3] O. Bälter, Strategies for Organising Email, 1997,
- [4] D.K. Barreau, Context as a Factor in Personal Information Management Systems, *Journal of the American Society for Information Science* 46 (5), 1995, pp. 327-339.
- [5] D.K. Barreau and B.A. Nardi, Finding and Reminding: File Organization from the Desktop, *SIGCHI Bulletin* 27 (3), 1995, pp. 39-43.
- [6] R. Boardman, Improving Tool Support for Personal Information Management, unpublished Doctoral Dissertation, Imperial College, 2004.
- [7] R. Boardman and M.A. Sasse, "Stuff Goes into the Computer and Doesn't Come Out" A Cross-tool Study of Personal Information Management, *Proceedings of the CHI'2004 Conference on Human Factors in Computing Systems*, Vienna, Austria, 2004.
- [8] R.E. Boyatzis, Transforming qualitative information: Thematic analysis and code development, Sage Publications, Thousand Oaks, California, USA, 1998.
- [9] P. Drucker, Landmarks of tomorrow, Harper Collins, New York, 1959.
- [10] N. Ducheneaut and V. Bellotti, E-mail as Habitat: An Exploration of Embedded Personal Information Management, *Interactions* 8 (5), 2001, pp. 30-38.
- [11] A. Edmunds and A. Morris, The problem of information overload in business organisations: a review of the literature, *International Journal of Information Management* 20 (1), 2000, pp. 17-28.
- [12] J. Faichney and R. Gonzalez, Goldleaf Hierarchical Document Browser, *Australian Computer Science Communications* 23 (5), 2001, pp. 13-20.
- [13] A.F. Farhoomand and D.H. Drury, Managerial Information Overload, *Communications of the ACM* 45 (10), 2002, pp. 127-131.
- [14] J. Gwizdka, Email Task Management Styles, *Proceedings of the CHI'2004 Conference on Human Factors in Computing Systems Extended Abstracts*, Vienna, Austria, 2004.
- [15] P. MacInnis, Desktop democracy, *Computing Canada* 29 (13), 2003, pp. 10.
- [16] W.E. Mackay, More than just a communication system: diversity in the use of electronic mail, *Proceedings of the CSCW'88 Conference on Computer-Supported Cooperative Work*, Portland, Oregon, USA, 1988.
- [17] T.W. Malone, How do people organize their desks? Implications for the design of office information systems, *ACM Transactions on Office Information Systems* 1 (1), 1983, pp. 99-112.
- [18] J. Nielsen, First Rule of Usability? Don't Listen to Users, 2001, <http://www.useit.com/alertbox/20010805.html>.
- [19] OneStat.com, Microsoft's Windows dominates the OS market on the web according to OneStat.com, 2004, http://www.onestat.com/html/aboutus_pressbox24.html.
- [20] L. Richards, Handling Qualitative Data: A Practical Guide, Sage Publications, London, UK, 2005.
- [21] D.L. Schacter, The Seven Sins of Memory: Insights From Psychology and Cognitive Neuroscience, *American Psychologist* 54 (3), 1999, pp. 182-203.
- [22] N. Schwarz, Self-Reports: How the Questions Shape the Answers, *American Psychologist* 54 (2), 1999, pp. 93-105.
- [23] N. Schwarz and D. Oyserman, Asking Questions About Behavior: Cognition, Communication, and Questionnaire Construction, *American Journal of Evaluation* 22 (2), 2001, pp. 127-161.
- [24] H.A. Simon, The future of information systems, *Annals of Operations Research* 71 (0), 1997, pp. 3-14.
- [25] J. Teevan, C. Alvarado, M.S. Ackerman and D.R. Karger, The Perfect Search Engine Is Not Enough: A Study of Orienteering Behaviour in Directed Search, *Proceedings of the CHI'2004 Conference on Human Factors in Computing Systems*, Vienna, Austria, 2004.
- [26] S. Whittaker and J. Hirschberg, The Character, Value, and Management of Personal Paper Archives, *ACM Transactions on Computer-Human Interaction* 8 (2), 2001, pp. 150-170.
- [27] S. Whittaker and C. Sidner, Email Overload: exploring personal information management of email, *Proceedings of the CHI'96 Conference on Human Factors in Computing Systems*, Vancouver, Canada, 1996.