Published in 'Contemporary Ergonomics 1990', Ed. E.J.Lovesey, Taylor & Francis 1990 (the proceedings of the Ergonomics Society 1990 Annual Conference, Leeds, UK)

ERGONOMIC ASPECTS OF COMPUTER SUPPORTED PERSONAL FILING SYSTEMS

PAUL WILSON

Computer Sciences Company Limited,
Computer Sciences House, Brunel Way,
Slough, SL1 1XL, UK. Tel: +44-753-73232
JANET: wilson@cs.nott.ac.uk HICOM: paul_wilson@uk.ac.lut.hicom

This paper reports on a long term empirical Personal Filing (PF) study which has already lasted over 8 years. It emphasises that PF represents a different problem set to Library and On-Line Database filing and retrieval, and identifies **speed** in filing activities as the critical factor in PF systems. The filing index schema and physical filing arrangements are described. Other issues discussed include types of search and different search circumstances, self-adjusting index terms, optimisation of the work/filing relationship, multiple users and document types, and archiving.

INTRODUCTION

In 1980, while working for The National Computing Centre's newly constituted Office Systems Division, my colleague, John Pritchard, and I visited the Amoco Production Company's Research Centre in Tulsa. There we saw our first working automated office filing system: all documents were allocated serial numbers, titles and keywords which were held in a computerised index; and the hardcopy documents were physically filed according to their numbers. Retrieval was done by searching the electronic index and then accessing the hardcopy. Our Amoco host, James G Steward, told us this system was very successful.

This and many other visits made John and I impatient to try out some Office Technology concepts. At that time, having no office systems ourselves, we were reduced to trying out office automation principles in manual mode. Our first such investigation was to replace our desk diaries with documents which were updated regularly in the typing pool (and subsequently on our own workstations) (Wilson, 1984). The second was to agree a schema for indexing and sharing our documents, and to apply it to our own personal files using card indexes. In the 8+ years since we first agreed the schema I have acquired over 4600 indexed documents (now in a computerised index), a wealth of practical experience about how to operate my filing system effectively, and a commitment to continue the exercise as a long term empirical investigation of personal filing in the electronic age. My overall objective is to provide practical feedback to product developers and system designers regarding the real day to day requirements of individuals using personal electronic filing systems. I now perceive the investigation to consist of three main phases:

1. The operation of a card-based index of hardcopy documents.
2. The operation of a computer-based index of hardcopy documents.
3. The operation of a computer-based index of document images held on optical disc.

The first phase was completed at the end of 1987, and I am now in my third year of phase 2. Phase 3 will consist of putting all my documents onto optical disc, and learning what its like to work with electronic documents on a day to day basis. I shall proceed to phase 3 just as soon as I have written up my experiences to date (Wilson, in progress), produced a prospectus for potential system sponsors and obtained the necessary equipment and systems. In the meantime I continue to retain core data (such as all documents input to the system, and details of all retrievals made each day) for ongoing analysis.

Filing systems are complex and have many facets. This paper, however, discusses just a few of those facets which have a clear Ergonomic element. First, though, I will briefly describe the general schema of the filing system I am using.

THE FILING SCHEMA

The basic filing schema I agreed with John Pritchard and some of our other colleagues, consisted of 3 major components:

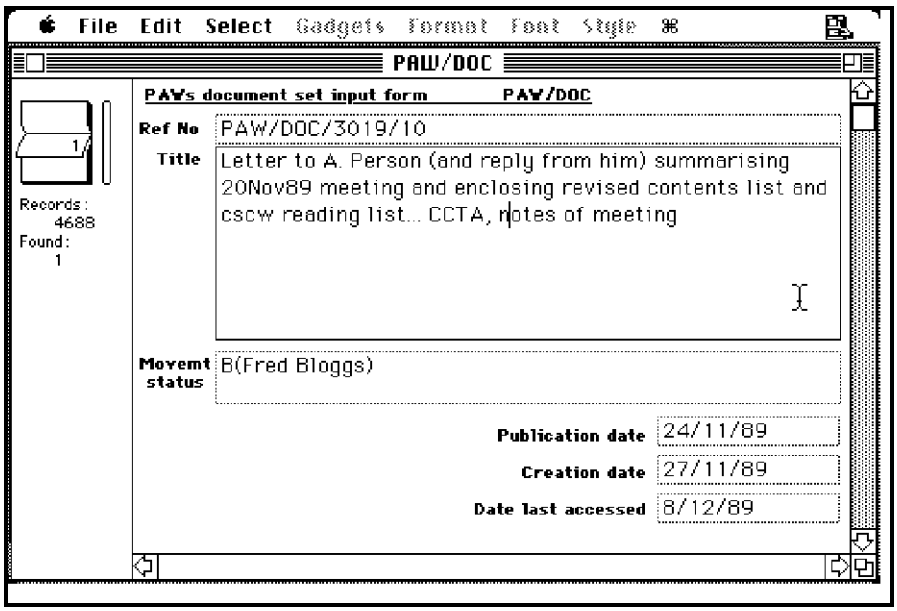
- A **Reference Number**, to be applied to each document and having four parts - an owner identifier (eg PAW for Paul Wilson), a set number (eg DOC), a serial number (eg 3010), and a sub-serial number (eg 01). So, a typical Reference No looks like this: PAW/DOC/3010/01. The purpose of the owner identifier is twofold; first to allow two or more people to each generate unique numbers, and secondly to enable references to other people's documents to be included in one's own index while being clear about their physical location. The object of the set identifier is to allow an individual owner to have several different physical sets of information. The purpose of the serial number is to enable new documents to be given the next number on the list i.e.the number signifies nothing other than the physical location of the document in the file. The purpose of the sub-serial number is to enable two or more documents to be kept physically together in a file if thought appropriate by the owner. I write Reference Nos in blue felt tip pen at the top right hand corner of documents.
- A **Title,** to be applied to each document, unlimited in length and the contents to be at the owners discretion (i.e. it could be different to the actual title on the document).
- **Keywords or phrases,** to be applied to each document, unlimited in number, specified entirely at the owner's discretion and added to the end of the title after three dots (eg... keyword)

In addition to these fundamental components, my computerised index also contains the following fields:

- **Movement status,** to record the location of items Borrowed by people, Archived, Lost etc.
- **Publication date,** to indicate when an individual item was first published.
- **Creation date,** to record when an item was catalogued into the filing system
- **Date last accessed,** to record when an item has last been retrieved. This enables items which have never been accessed, or least recently accessed, to be identified as most suitable for archiving.

An example of a document record incorporating these components is shown in Figure 1.

Figure 1. The filing system input screen illustrating all the record fields



PHYSICAL REQUIREMENTS

The object of filing is to be able to find and use relevant information quickly and easily. The associated physical arrangements have as much impact on its effectiveness as the logical arrangements. So, I file my documents in an upright cabinet positioned to form a right angle with the left side of my desk. Hence, to access files I simply turn in my swivel chair (see figure 2).

Figure 2 Accessing files from the cabinet

With my computer (originally my card index box) also on the left side of my desk I try to reduce the physical movement required to input new items and retrieve existing items, to a minimum. Six office moves in 8 years has given me plenty of scope to try out a variety of layouts:

whenever I change offices I always aim to have my upright filing cabinet, bookcases and drawers all around me and accessible via minimum movement in the swivel chair.

The type of filing cabinet also has a critical impact on filing effectiveness. Four drawer cabinets, so prevalent in modern offices, are simply not as useful as upright cabinets. They do not hold enough material, have to be opened each time you want to access them, and are difficult to access from the seated position. Upright cabinets have few of these disadvantages, though, for maximum effectiveness, you need one with a front cover that slides up, not with doors that open.

Although a variety of physical filing systems are available for upright cabinets I have always used hanging folders and been very satisfied. Hanging folders usually have crystal tabs in which to display index information. I originally planned to use the tabs to display the Reference No of the first item in each hanging folder, but soon realised this was impractical since the contents are never static (due to inserting items using the Sub-serial No, adding to items using a stapler, or removing items for archiving). It is not necessary anyway since many Reference Nos are visible when the cabinet is open and that, combined with the mental map one builds up of the locations of Reference Nos, supports the initial foray into the cabinet when searching. From then on, knowing that items are filed in Serial No order, it is a matter of trying a few more items until the right one is hit upon. It is not perfect but I am satisfied with this physical search method.

SPEED - THE CRITICAL FACTOR

Enormous amounts of work have been done on filing and retrieval - but much of it seems to have been aimed at storing and retrieving literature in libraries and on-line databases. Typical priorities seem to be issues such as the dilemma between **Recall** (the number of items retrieved), and **Precision** (the percentage of relevant items retrieved); and the difficulties users have in formulating their search enquiry (see, for example, ACM Transactions on Office Information Systems, July 1989; or Sparck Jones, 1974; or Bärtschi, 1985). Useful studies of personal filing do exist (eg. Stibic, 1980; Moon, 1988), however few seem to address the filing of all types of document, nor the practical problems of doing so as a secondary, very low priority, work activity.

This is unfortunate because there are a completely different set of problems in personal filing activities. For example, there is no doubt in my mind that there is one overriding critical factor in personal filing: speed! The reason is one of which most knowledge workers have bitter experience: filing is the bane of office work. It has little immediate payback and consequently gains little priority. Its also hard work to do effectively, and it gets harder as we deal with more and more paper in the electronic office (!).

Speed is important in every aspect of personal filing: inputting new items, searching the index, retrieving documents, putting documents away, and archiving documents. Seconds are important: a 5 second wait for search results can seem like a long time when you're busy. Cataloguing a new item might just not seem feasible if it takes more than a couple of minutes when you have a document to prepare for a meeting in half an hour.

The need to increase speed has affected many of my design decisions. For example, when I computerised my index I had the opportunity to add a great many more fields to the item record: but I stuck with the minimum I could get away with - extra records simply make for extra work. I could have made Keywords a separate field, but that would have added time to the creation of search enquiries - so I left them tagged onto the end of the Title field enabling me to search only on the Title field most of the time.

Naturally, I get the computer to do as much of the work as possible when I input a new item: the machine generates the Creation Date and the first part of the Reference No (PAW/DOC/), and I finish the Reference No and create the Title/Keywords and, where appropriate, Publication Date. Another way I speed things up is to use a macro package (Tempo) in conjunction with my database software (FileMaker Plus) so that single keystrokes take me directly to the Search screen and New Item screen; and I use Multifinder on my Macintosh to enable me to keep the index present at the bottom of my screen whatever else I'm doing. That way its never more than a mouseclick away throughout the working day.

Speed also has an important part to play in the discussion about just what should be filed and what shouldn't. I file just about everything because I want to investigate the worst case (i.e greatest filing load) scenario, and also because I want to build up a complete office document test set (the homogeneous nature of office documents means that library test sets cannot reflect the office filing problem. Several library test sets are in existence enabling much work to be done on library retrieval problems; the lack of office document test sets has almost certainly constrained research in this area [Wilson, 1986]). Naturally filing everything takes more time, so filing selectively might reduce the work load. On the other hand filing everything means I don't have to think about what to file, nor about whether something is filed or not when I do a search. So it may be that the time I lose in filing everything is not actually a loss in practice. I don't know the answer to this question yet. Its certainly a complex issue and, like a lot of other issues relating to the paramount factor of speed, could do with some intensive research.

PRACTICAL SEARCH AND RETRIEVAL ACTIVITIES

For the personal filing system there is no such thing as a standard search and retrieval activity. There are a variety of **types** of search, and a number of different **circumstances** under which searches take place. Search types include:

- the **Familiar Item** search for a specific document that you know well, are familiar with what search terms to use to find it, and may even know roughly where it sits in the physical file.
- the **Long Lost Friend** search for a document that you're pretty sure you have but can't remember too much about how it is indexed or where it might be in the physical file. This and the Familiar Item type of search make up the vast majority of search instances.
- The **Shot in the Dark** search to find if you've got some material on a subject. You have to do this sometimes because, after a while, you just don't know what's in your filing system.
- The **Literature Search** for everything you've got on a particular subject. You may have to try a number of different enquiries to get a result here, and more often than not you'll be selecting items from the hit lists the search throws up.

Search circumstances include:

- The **Start Work** circumstance where there's little pressure and you can concentrate on finding the appropriate material before getting down to a particular work activity.
- The **Mid Work** circumstance where you are in the middle of doing something and you suddenly need some information. This circumstance is slightly different from the Start Work circumstance because your mind is somewhat preoccupied and the search and retrieval process is in effect an interruption. In this circumstance you may also be using another application on the computer, necessitating a switch to the filing application. Before I got Multifinder on the Macintosh this was a real problem because I had to quit out of, say, the word processing application, do the search and then open up the word processing program again - which was a very slow and frustrating process. Multi-tasking systems, however, eliminate this problem.
- The **Visitor** circumstance when you need to get at a document while you are talking to someone at your desk. This requires either that you interrupt the conversation and explain you are just going to search for a document, or that you continue to talk animatedly while you do an index search and then look for the physical document.
- The **Phone Call** circumstance in which you try and find a document while in the middle of a phone call, still clutching the phone handset, without breaking the flow of the conversation. This and the Visitor circumstance are particularly exacting tests of filing system ergonomics, demanding high levels of speed and ease of use characteristics in both the indexing system and the physical filing arrangements. Designs that cater for the Phone Call circumstance would probably meet most other search and retrieval usability requirements.

SELF-ADJUSTING INDEX TERMS

After you use a personal filing system for a few years you begin to realise that your language is changing. The terms you were using now seem inappropriate; and you find you've stopped using some full length index terms and replaced them by abbreviations or acronyms.

This shouldn't surprise us as our day to day language is constantly changing and, naturally, this is reflected in our filing systems. Unfortunately it makes the search and retrieval process more difficult. For example, when I first started work on email systems I used the keyword phrase 'electronic mail'. Some months later, when I had become immersed in the topic I adopted the more general term 'mailbox system' instead. Now, several years later, I seem to have abandoned the latter term and often simply use 'email' (these changes, by the way, reflected not only language changes, but also work priorities: while working on the electronic mail topic I was preoccupied in getting my classifications right, but now I am much more relaxed about it).

Another example is when I started work on trying to organise some follow up activity on a research programme I was involved with: I started filing material using the term 'CSCW programme'. After a while a group known as 'FDG' was formed to undertake this work, so I would have preferred to use FDG as the master keyword. Unfortunately, however, to ensure I would retrieve a full listing of all relevant items each time I did a search I had to continue using the original 'CSCW Programme' term. Now that the work is coming to fruition a new term - 'CSCW Foundation' - is becoming more appropriate.

One partial solution to such problems might be to use a synonym facility so that whenever you recognise there has been a change in your vocabulary you can tell the system to search for the old word when you specify the new word. Unfortunately the system I'm using doesn't have such a facility so I can't report on its usefulness. However, it might not actually solve the problem since I rarely specify whole words in the search process. For example, in searching for 'mailbox system' I would normally specify something like 'mai sys' (putting in whole terms would just slow things down). This would fail to activate a synonym substitution of 'electronic mail' for 'mailbox system'.

Instead, some sort of knowledge based system is required which monitors the language being used in the new records being input, in the enquiry terms being specified for searches, and in the records being selected for retrieval as a result of each search. Four important components of the Language Monitor would be:

- Noting what terms tend to appear together in the title/keyword field in each new record.
- Matching the shortened search terms to the full terms they 'hit'.
- Relating search terms to the terms that appear in records selected for physical retrieval.
- Occasionally requesting the user to confirm, or otherwise, the conclusions it is coming to about the relationship between terms, and the user's changing language.

In its turn I can see the Language Monitor being an important component of some sort of future Filing System Agent which would interface between the user and the system in several different respects.

OPTIMISING THE WORK/FILING RELATIONSHIP

Filing is not some independent activity with its own rationale for existence. We file to support our work; so our work and filing activities need close and careful integration (Lansdale, 1988). There are many facets to this complex issue - not least being that everyone has different ways of working (Malone, 1983). However, some of the issues that are important **for me** are:

- Unfiled paper complicates and distracts. It diverts efforts from real work tasks by continuously requiring assessment, sorting and tidying, and by demotivating the individual (Oliver, 1989). I'm more efficient if I read and file paper when I receive it. If its something I can't deal with straight away its better if I file it and put an action item on my 'to do' list - along with a file reference number so I can retrieve the item without having to search the index first.
- Knowing what documents are relevant for particular work tasks helps me decide how to file them: for example, I tend to staple committee documents together (rather than filing each independently) because I don't often use them much between committee meetings. This way its easy and quick to retrieve all the relevant documents on the way out to a meeting. Such

decisions about what to file with what, or what to file and index separately have to be made constantly and have a significant impact on the work process.

- I do quite a lot of work away from the office and at home, so the filing system has to be usable in those circumstances as well as in the office. In particular, I need to be able to file new material and to access filed items. At present I can only do the first of these (by taking my Macintosh home, or by noting the next serial Reference Number and creating new index entries on paper which I input at a later date). However, being able to retrieve the actual filed items when I'm away will have to wait till I have an effective portable optical disc system.

OTHER ISSUES

Two other issues I planned to discuss but am unable to because I have used up the six allowable pages, are Multiple Users and Document Types, and Archiving.

Multiple document types, and multiple users (in a shared filing environment) demand high degrees of flexibility from the filing system.

Archiving, to release physical filing space, is an unavoidable and time-consuming aspect of filing. Until paper is eliminated entirely (not for a long, long time yet) it will be important to optimise every aspect of the archiving process.

REFERENCES

ACM special issue on Research and Development in Information Retrieval, July 1989, ACM Transactions on Office Information Systems, Vol 7 No 3.

Bärtschi, M., May 1985, An overview of Information Retrieval subjects, Computer, IEEE.

Lansdale, M., 1988, The psychology of personal information management, Applied Ergonomics, 19.1 (55-66).

Malone, T., 1983, How do people organise their desks? Implications for the design of office automation systems, ACM Transactions on Office Information Systems, 1 (99-112).

Moon, C., 1988, Computerised Personal Information Systems for Research Scientists, International Journal of Information Management, 8 (265-273).

Oliver, N., November 1989, Jitters in the office, Management Today (142-149).

Sparck Jones K., 1974, Automatic Indexing, Journal of documentation, Vol 30 No 4 (reprinted in Progress in Documentation: some Themes and Topics, ASLIB, 1981)

Stibic, V., 1980, Personal Documentation for Professionals, North Holland.

Wilson, P., 1984, Towards the electronic pocket diary, Design Studies, Vol 5 No 2, 98-105

Wilson, P., 1986, Introducing Electronic Filing, NCC Publications.

Wilson, P., In progress, Personal filing in the electronic age - A Long Term, Empirical Investigation