Information Workers and their Personal Information Management: a Literature Review

A.A.J. (Jos) van Helvoort

The Hague University of Applied Sciences, Academy of ICT&Media, Post Box 13336, 2501 EH The Hague, The Netherlands. Email: a.a.j.vanhelvoort@hhs.nl

Abstract: *The research described in this paper provides insights into tools and methods which are used by professional information workers to keep and to manage their personal information. A literature study was carried out on 23 scholarly papers and articles, retrieved from the ACM Digital Library and Library and Information Science Abstracts (LISA). The research questions were:*

- *How do information workers keep and manage their information sources?*
- *What aims do they have when building personal information collections?*
- *What problems do they experience with the use and management of their personal collections?*

The main conclusion from the literature is that professional information workers use different tools and approaches for personal information management, depending on their personal style, the types of information in their collections and the devices which they use for retrieval. The main problem that they experience is that of information fragmentation over different collections and different devices. These findings can provide input for improvement of information literacy curricula in Higher Education. It has been remarked that scholarly research and literature on Personal Information Management do not pay a lot of attention to the keeping and management of (bibliographic) data from external documentation. How people process the information from those sources and how this stimulates their personal learning, is completely overlooked.

Keywords: *Personal Information Management; information literacy; higher education.*

Introduction

The research reported in this paper has been conducted in the context of my research activities in Information Literacy. The ACRL Information Literacy Competency Standards for Higher Education mention as performance indicator 2.5: "the information literate student extracts, records, and manages the information and its sources" (ACRL, 2000). The aim of the research presented here is to construct an overview of tools and methods which are used by professional information workers to keep and to manage their information. This overview might suggest directions for the improvement of Information Literacy Curricula in Higher Education (Jones 2007, p. 457).

Research Approach

Research Questions

The main question of the research was: What kind of tools and approaches do information workers use to keep and to manage their information sources?

Subsequent questions were on the aims they have when keeping information and the problems they experience.

These research questions belong to the field of Personal Information Management (PIM), a research field that includes users' activities to acquire or to create information items and the retrieval, use and distribution of them (Jones 2007, p. 457). However, in the context of the ACRL performance indicator 2.5 the present research is focussed on storing, organising and maintaining information with the aim of creating possibilities to locate it for later use.

Methods

The research was conducted as a literature study. The first step was desk research of relevant scholarly literature. The second step was the analysis of its content. For the content analysis the results of the desk research were grouped in themes (for instance "Types of collections that people create") and for each theme the main findings are presented.

Search Strategy

The first document that was read on the topic PIM was Jones' chapter in the Annual Review of Information Science and Technology (2007). In the next step of the orientation with Google Scholar it was observed that a lot of scholarly literature on PIM consists of conference contributions from which the papers are published in the ACM Digital Library. A search in the ACM Digital Library with the *discovered terms* "Personal Information Management" and "User Studies" resulted in a hit list with fourteen references (27 November 2010). After further reading, one of them appeared not to be relevant to the research question. Ten of the thirteen relevant documents were full text available and were used for this study.

Because the ACM Digital Library focuses on Computer Science the search was extended with a query in the main source for Library and Information Science LISA on 28 November. Search terms were "personal information management" AND NOT "book review", limited to peer reviewed journals in the English language. This query resulted in twenty one hits. After first reading it appeared that sixteen of them were relevant for my own research question. Thirteen titles were full text available and were used for the research. There was no overlap with the titles from the ACM Digital Library.

Findings

Types of Collections that People Create

The concept of PIM refers to information management activities in various domains: email messages, contact lists, files and documents in a "my documents" folder, references to web pages and tangible documents like photocopies, prints, handwritten notes and books. Jones refers to those domains as "Personal Information Collections" which together form ones "Personal Space of Information" (2007, p. 462). The elements of a collection mostly share a "particular technological format and are accessed through a particular application" (Boardman 2004, cited by Jones 2007, p. 462). Most collections described in the literature are local files, email and bookmarks. Remarkable is that in the Computer Science literature no attention is paid to the use of bibliographic software (for instance Endnote, Refworks or Zotero). From the titles which were retrieved from LISA only McGeachin (2004) and Yu (2008) pay explicit attention to bibliographic citation management by scholars.

How People Organise their Personal Information Collections

Indratmo and Vassileva (2008) distinguish five different approaches as to how people organise their personal information: hierarchical, flat, linear, spatial and networked. Besides these five approaches a sixth approach can be distinguished which uses a minimum of organisation and is for that reason very popular to end users of information: full text search. All these six approaches will be discussed here in more detail while remarks and insights from different authors will be added to the elaboration.

Hierarchical approach

The hierarchical organisation of information items (for instance in folders and subfolders) is most commonly used by information workers. People are very familiar with it and therefore the process of organising information in a hierarchical way is very intuitive (Indratmo & Vassileva 2008, p. 2). The drawbacks are the problems in creating a consistent structure and difficulties in choosing which category one should file an item (Bergman et al. 2008[a], p. 235; Bergman et al. 2008[b], p. 3; Indratmo & Vassileva 2008, p. 3; Jones 2007, p. 473). Nevertheless, the hierarchical approach seems to be a good solution if the information user needs a complete overview of all the relevant documents (high recall).

Indratmo and Vassileva refer to this situation in the context of project management (2008, p. 10). In practice the use of hierarchical folders is quite common for information that belongs to specific tasks when it is easy to determine to which task a document belongs (one folder for each task). However, the hierarchical approach hinders retrieval by means of mobile devices such as smartphones, due to their small screens. For those devices, full text search or sending relevant documents to the working space seem to be better solutions (Glória et al. 2010).

Flat approach

With the "flat approach" Indratmo and Vassileva refer to the practice of labelling or tagging information items (Indratmo & Vassileva 2008, p. 4 ff.). The most striking advantage of the flat organisation method is that users can classify an information item into multiple categories simply by assigning multiple tags to the item (Ibid. p. 4). The drawback is of course the risk of "noise" in the retrieval phase. Indratmo and Vassileva therefore emphasise that the flat approach is particularly suitable for more general collections of information (2008, p. 10). The flat approach is best known in different tools for online bookmarking (for instance Delicious).

Linear approach

With the linear approach of storing and retrieving information people use the function of an information system to sort the output of a query or an information collection, for instance chronologically or alphabetically. This approach is well known in email clients,where one can sort messages under the sender ("from") or a date ("received") or in applications to browse photos ("date created"). It is often used as a second approach within a folder or a set of information items with a common tag (Boardman & Sasse 2004, p. 587). This technique is also meant by Indratmo and Vassileva when they plead for providing multiple visualisations (Indratmo & Vassileva 2008, p. 11).

The common habit of storing books or prints of journal articles alphabetically, under author names, is another example of the linear approach.

Spatial approach

A spatial approach to organising information items means that frequently or recently used items are presented for quick access on, for instance, the user's computer desktop. Indratmo and Vassileva describe a tool for personal information management that helps users to arrange their information items on a special desktop without using folders (TMC, Time-Machine Computing. Indratmo & Vassileva 2008, p. 7) but more frequently this approach will be used by information workers in their document applications with a sidebar that allows them to find a shortcut to the last used documents. The Inbox of an email client is also often used as a spatial tool in which the current work can be found.

The spatial approach is often combined with "past interaction" (Bergman et al. 2003, p. 875-876; Indratmo & Vassileva 2008, p. 12). The electronic traces that people leave on the internet and on their local systems are, in this approach, used for quick access to their documents. The best known example of this approach is without doubt the history button of a web browser. "Past interaction" is also the principle that is used for access to recently used documents in text editor applications, as discussed earlier.

A spatial approach could also be a solution for archiving purposes. Keeping apart the most valuable digital information makes it easier to manage it for availability in the long term (Cushing 2010, pp. 305-307). However, it appears that archiving of personal information still lacks enough attention in the research and literature on PIM (Ibid.; Williams et al. 2009, p. 340).

Network approach

This approach is a rather sophisticated method of organising information. It works with the linking principle that is also used to organise the World Wide Web. Creating links between information items (even between items in different information collections) is quite easy but in the retrieval phase it is hard to navigate and to find your way in the network (Indratmo & Vassileva 2008, p. 8). Just as on

the World Wide Web it is easy to find one relevant item but finding all relevant items ("high recall") is a complicated task.

Minimal structure: full text search

With full text search in their personal information collections people use characteristic phrases or author names to refind the information items. Search technology potentially solves the problems that information users have with organising their information items in hierarchical folders (Bergman et al. 2008[b], pp. 3-4). However, empirical studies show that, for personal information management, people still prefer a hierarchical folder and navigation method over search (Barreau 2008, p. 312; Bergman et al. 2008[b], p. 16). Bergman et al. suggest that retrieval behaviour in the context of personal information management differs from retrieval behaviour on the World Wide Web because "in PIM, the same user both organizes and retrieves the information" (Ibid. p. 20). Moreover, they argue that the organisation of information items in folders functions as "sense-making" (see also Jones 2007, p. 487).

For full text search, the role of memory is essential. People have to remember characteristic words from the text, title or other bibliographic data. Sometimes this is a problem. A lot of the conference literature is on experiments to improve integrated desktop search with autobiographic information that people remember (Dumais et al. 2003; Elsweiler et al. 2007; Gonçalves & Jorge 2008) or associations with events and persons (Chau et al. 2008). These kinds of tools make it possible to get an answer to a query like: "the document that I received from the woman whom I met at the conference in Budapest". Bergman et al. confirm the importance of this kind of "subjective attributes" of information items in personal information management (Bergman et al. 2003, p. 873; Bergman et al 2008a, p. 236). However, Elsweiler, Baillie and Ruthven found that, at least for email, this kind of information is less often remembered than the topic of the message or the reason it was sent (Elsweiler et al. 2008, p. 31). In a more recent research article they concluded that this might be true but that it does not mean that remembering the topics of email messages also leads to better performance of refinding tasks. They conclude that finding tools in email clients should provide better support for semantic retrieval of information (Elsweiler et al. 2009, p. 24).

Different approaches combined

Boardman and Sasse discovered that people use different organising strategies for different collections. More effort is invested in organising their document files than in organising email or bookmarks because of the effort they have initially spent on creating their documents. Email collections are typically retrieved with the linear approach (Boardman & Sasse 2004, p. 589) and particularly "no filers" have a good memory for the information in their email collection (Elsweiler et al. 2008, p. 22). More in general the authors of the previously mentioned article plead for flexibility in refinding tools when designing interfaces for email clients. It appears that for the retrieval of information most people prefer a mix of different tools and approaches (Elsweiler et al. 2008; but also Elsweiler et al. 2007, p. 934). An example of this is a hierarchical search combined with alphabetical or rather chronological presentation of the results (for instance: recently used documents on the top of the list).

Descriptions that People Use to Keep their Individual Information Items

Boardman and Sasse (2004, p. 586) classified folder names that people used for the folders with their email, files and bookmarks. They found names for projects (short term activities), document properties (for instance 'letters'), roles (long term activities, for instance 'teaching'), contacts and topics/interests. In their research on the way in which people organise their information items, Bergman et al. 2008[a] (p. 239) found that most computer users tend to use project names for their folders more than file properties (for instance their technical 'format'). This is consistent with Kwasnik's findings (1989, cited by Williams et al. 2009, p. 352). However, it is contrary to Barreau's conclusion that for electronic documents the managers she interviewed most often used the document attribute "Form" (which includes types of documents as well as file formats) for folder names. But it has to be said that she only interviewed four managers and that those managers recognised that

"what is easy and convenient for them may not be best practice for their organization as a whole" (Barreau 2008, p. 316).

Aims that People have when Building Personal Information Collections

During the performance of a task people seek information to address an immediate information need. They keep it in their systems, for instance in their bookmark collection or in their email client as a self-addressed email message for future use (Bruce 2005). However, the collecting of information is not always purposeful. Coughlan and Johnson found how creative practitioners collected interesting materials and ideas which came along more or less accidentally and which were not really related to the projects they were currently working on (Coughlan & Johnson 2009, pp. 7-9).

In daily life people also encounter information incidentally (for instance by listening to a radio broadcast or talking with colleagues) and they match it to "anticipated needs" (Bruce 2005; Jones 2007, pp. 472-473). That is why Bruce (2005) remarked that enhancing "sensitivity and appropriate responses to *personal, anticipated information need* is […] a key component of information literacy".

"Keeping decisions" are fundamental to personal information management (Jones 2004). Keeping too many items may distract a person's attention from information that really is useful. On the other hand, the strategy of keeping nothing and trusting that one will retrieve it again on, for instance, the World Wide Web at the moment that one needs it, might lead to the situation that one forgets its existence entirely (Ibid.). Nevertheless, Barreau also found that managers trust more in searching again than in keeping bookmarks for information on the Web (Barreau 2008, p. 316).

The management of personal information collections is also a component of someone's learning process (Bergman et al. 2003, p. 872). Information items are read, interpreted and cognitively processed. As mentioned earlier, some authors believe that the majority of information users still prefer organising their information items in a hierarchical folder system because the classification of the items also functions as "sense making" (for instance Bergman et al. 2008[b], p. 20).

A third reason why people store information items is to remind them of a task that they have to perform (Boardman & Sasse 2004, p. 583; Elsweiler et al. 2007, p. 925). Particularly the email inbox is often used as a "to do" list (Barreau 2008, p. 310).

Yu (2008) noticed a remarkable alternative possible aim of gathering information by publishing scholars. The model that he proposes not only encourages scholars to create a personal information collection for citation purposes but also stimulates them to share their knowledge sets with other scholars. For the sharing scholar this has the advantage that it increases his visibility in the scholar world.

Problems Experienced

Fragmentation of information

According to Jones (2007, pp. 453-456) one of the biggest problems that people experience with personal information management is the fragmentation of information. As described earlier, information items are stored in a variety of information collections. Indratmo and Vassileva (2008, p. 10) give the example of information items related to a project but that are stored in a file system, an email client and a calendar. This enumeration can easily be expanded with, for instance, financial information from an administrative system and bibliographic records from a citation manager.

Bergman et al. propose a couple of improvements on interface design to enable computer users to organise their information items in a project based way (Bergman et al. 2008[a], pp. 241-243). However, the most commonly used approach to integrate information items from different collections is without doubt "integration through search" with "cross-format" desktop search facilities, for instance Google Desktop, or Spotlight in the Mac operating system. But this technological approach is not a solution for all information management problems, for instance not for the problem of multiple versions of documents (Jones 2007, p. 487). Moreover, as discussed earlier, search engine technology has the disadvantage that it presents access to individual documents and not to sets of documents that relate to each other.

Fragmentation of information not only occurs across different information collections but more and more also across various devices such as desktops, laptops, smart phones and personal digital assistants (Jones 2007, p. 475). Storing information items "in the cloud" could contribute to a solution for this different devices problem, although Dearman and Pierce report that academic professionals are more willing to use such web-based services than employees in industry (2008, p. 773).

People have to realise that they have an information need

Jones (2007, p. 468) emphasises that there is an important step preceding the retrieval of information from the personal space of information. He calls this "Remembering To Look". This step is comparable to "Recognising an Information Need". When people do not realise that they have an information need for the task at hand they will indeed forget to look up the information in their information space. Indratmo and Vassileva discuss a solution for this problem by the use of "future" or "planned interaction" records that remind information users of "potentially relevant documents" (Indratmo & Vassileva 2008, p. 12; but also Chernov 2008). Jones, however, also made a remarkable statement when he observed that keeping information also enables people to remember to look for it at a later time when the need arises (Jones 2007, p. 477). He even argues that putting an information item in a folder enables the cognitive processing of the content of the item (Ibid. p. 487). However, this is in contrast with Elsweiler and others' earlier mentioned conclusion that "no filers" have a good memory for the information in their email collections.

Keeping decisions

In the previous section I discussed the problems that information users often have with "keeping decisions" (Jones 2004). To solve the keeping problem Bergman et al. suggest a sorting tool for information retrieval that puts less used items at the bottom of the result list with a different appearance (for instance a smaller font and in a grey background) (Bergman et al. 2008[a], p. 243). This solution could also be used to present older versions of a document. As Bergman et al. argue, these design solutions combine the advantages of the "keep" option (the information item may be accessed in case it is unexpectedly needed in the future) with the advantages of the "delete" option (the information item no longer distracts the user's attention when he is looking for relevant information).

Discussion and Conclusions

The literature research described in this contribution was restricted to scholarly papers and articles from two different sources: the ACM Digital Library and LISA. This could be regarded as a limitation of the research but I do not have the impression that trends or developments in PIM are missed.

The research was started with the expectation that findings could contribute to the improvement of Information Literacy Education. In that context the results were a bit disappointing. Most of the literature that has been studied was on the management of files and emails. Only two articles discussed the management of (scholarly) literature and no attention at all was given to the way in which professionals extract that information. Only Williams et al. (2009, p. 350) remarked that this cognitive activity is indeed largely overlooked. The different contributions of Bergman et al. to PIM research confirm that the management of personal information stimulates one's learning process but it is not made very clear how this happens. This leads me to the conclusion that cognitive processing of information deserves more attention in PIM research. Potential research questions for this are:

- Which tools (for instance local bookmarks, online bookmarks or personal notes) do people use to manage references to information items that they have encountered and evaluate as possibly useful for future use and
- How do they express the content of these references (for instance with subject labels or summaries)?

Another conclusion from the literature is that information workers have a lot of problems with the fragmentation of information, over different collections as well as over different devices. It could be that students do not experience those difficulties to the same extent because they often work on fixed projects within a restricted period. However, over time they will probably be confronted with the same problems. Information literacy education should anticipate these problems by stimulating an

accurate attitude in document management. The same is true for the promotion of an attitude to anticipate the value of information that seems not to be valuable at that moment but that could be useful in the future. In that sense the development of PIM skills contributes to the education of life long learners.

The third conclusion from the literature is that information workers manage their information items in different ways. There is no one best approach. The preferred method(s) for organising someone's personal collections depend on personal style, the type of information that should be managed and the types of devices that are used for retrieval. The literature indicates that for computer files and emails project folders are a good solution because these items are often strongly related to the task that is being processed. But for more general documentation (not so strongly related to the task that is being executed at that moment) a flat approach might be preferred. However, when items should be retrieved with a mobile device it seems that "search" is the "smartest" way. This observation leads to my conclusion that also for information literacy curricula in Higher Education, it should be an aim to let students discover which approach(es) suits them best in different situations.

Acknowledgements

My acknowledgments go to two colleagues. The first one is Wouter Gerritsma from Wageningen UR who drew my attention to the topic PIM and William Jones' chapter on PIM in the ARIST. Peter Becker is acknowledged because he was so very kind to give his comments on the first version of this paper. They both made clear that also in Personal Information Management collaboration between people is a key factor for success.

References

ACRL. (2000). *Information literacy competency standards for higher education.* Chicago: American Library Association. Retrieved Nov. 25, 2010 from http://www.ala.org/ala/mgrps/divs/acrl/standards/standards.pdf.

Barreau, D. (2008). The persistence of behavior and form in the organization of personal information. *Journal of the American Society for Information Science and Technology, 59* (2), 307-317.

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

Bergman, O., Beyth-Marom, R., & Nachmias, R. (2008a). The user-subjective approach to personal information management systems design: evidence and implementations. *Journal of the American Society for Information Science and Technology, 59* (2), 235-246.

Bergman, O., Beyth-Marom, R., Nachmias, R., Gradovitch, N., & Whittaker, S. (2008b). Improved search engines and navigation preference in personal information management. *ACM Transactions on Information Systems, 26* (4, 20), 1-24.

Boardman, R. (2004). *Improving tool support for personal information management*. London: Imperial College.

Boardman, R. & Sasse, M. (2004). Stuff goes into the computer and doesn't come out: a cross-tool study of personal information management. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems.* Retrieved November 27, 2010 from http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.115.3372&rep=rep1&type=pdf.

Bruce, H. (2005). Personal, anticipated information need. *Information Research, 10* (3). Retrieved December 14, 2010 from http://informationr.net/ir/10-3/paper232.html.

Chau, D., Myers, B., & Faulring, A. (2008). What to do when search fails: finding information by association. *Proceeding of the Twenty-Sixth Annual SIGCHI Conference on Human Factors in Computing Systems.*

Chernov, S. (2008). Task detection for activity-based desktop search. *Proceedings of the 31st Annual International ACM SIGIR Conference on Research and Development in Information Retrieval.*

Coughlan, T. & Johnson, P. (2009). Designing personal information management systems for creative practitioners. *Journal of Digital Information, 10* (5). Retrieved December 1, 2010 from http://journals.tdl.org/jodi.

Cushing, A. (2010). Highlighting the archives perspective in the personal digital archiving discussion. *Library Hi Tech, 28* (2), 301-312.

Dearman, D. & Pierce, J. (2008). It's on my other computer!: computing with multiple devices. *Proceedings of the Twenty-Sixth Annual SIGCHI Conference on Human Factors in Computing Systems.*

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

Elsweiler, D., Baillie, M., & Ruthven, I. (2008). Exploring memory in email refinding. *ACM Transactions on Information Systems, 26* (4, 21), 1-36.

Elsweiler, D., Baillie, M., & Ruthven, I. (2009). On understanding the relationship between recollection and refinding. *Journal of Digital Information, 10* (5). Retrieved December 1, 2010 from http://journals.tdl.org/jodi.

Elsweiler, D., Ruthven, I., & Jones, C. (2007). Towards memory supporting personal information management tools. *Journal of the American Society for Information Science and Technology, 58* (7), 924-946.

Glória, H., Guerreiro, T., & Gonçalves, D. (2010). SmarterPhone: supporting meetings in a mobile context. *Proceedings of the 6th Nordic Conference on Human-Computer Interaction: Extending Boundaries.*

Gonçalves, D. & Jorge, J. (2008). In search of personal information: narrative-based interfaces. *Proceedings of the 13th International Conference on Intelligent User Interfaces.*

Indratmo, I. & Vassileva, J. (2008). A review of organizational structures of personal information management. *Journal of Digital Information, 9* (26).

Jones, W. (2004). Finders, keepers? The present and future perfect in support of personal information management. *First Monday, 9* (3). Retrieved November 28, 2010 from http://firstmonday.org/htbin/cgiwrap/bin/ojs/index.php/fm/article/view/1123/1043.

Jones, W. (2007). Personal Information Management. In: Blaise Cronin (Ed.), *Annual Review of Information Science and Technology* (pp. 453-504). Medford NJ: Information Today.

McGeachin, R. (2004). The impact of electronic bibliographic databases and electronic journal articles on the scholar's information-seeking behavior and personal collection of "reprints". *Science & Technology Libraries, 25* (1/2), 127-137.

Williams, P., John, J., & Rowland, I. (2009). The personal curation of digital objects: a lifecycle approach. *Aslib Proceedings: New Information Perspectives, 61* (4), 340-363.

Yu, S. (2008). Design of a model of publication sharing and harvesting. *The Electronic Library, 26* (4), 582-593.