Searching and Archiving: Exploring Online Search Behaviors of Researchers

Dhaval Vyas[1], Spencer de Groot[2], and Gerrit C. van der Veer[1,3]

[1] HMI Group, CTIT, Twente Universiteit, The Netherlands
d.m.vyas@cs.utwente.nl
[2] User-Centered Design Group, Elsevier Science, The Netherlands
s.groot@elsevier.com
[3] School of Computer Science, Open Universiteit, The Netherlands
gvv@ou.nl

Abstract. Searching for relevant peer-reviewed material is an integral part of corporate and academic researchers. Researchers collect huge amount of information over the years and sometimes struggle organizing it. Based on a study with 30 academic researchers, we explore, in combination, different searching and archiving activities of document-based information. Based on our results we provide some implications for design.

Keywords: HCI, Searching, Archiving, Design.

1 Introduction

Searching and archiving can be thought of as the two sides of online information seeking behavior, especially in the case of academic and corporate researchers. In order to support their professional activities and to be at the top of their research field researchers frequently search and archive document-based information from different sources. One of the major problems with this is that researchers collect more information than they can immediately use. In the field of HCI, most research on archiving and personal information management (PIM) [e.g. 1, 2, 6] alienate the searching activities from archiving. In this paper, we focus on understanding different searching and archiving activities of academic researchers. Our goal is to develop a holistic understanding of searching and archiving processes that is motivated towards designing efficient and effective technological support. Although several online search engines to a certain extent provide support for archiving (e.g. linking the documents to reference managers like EndNote), we believe that it is of paramount importance to understand what searching and archiving activities *mean* to researchers. We aim to understand what kind of information researchers look for when they search, how they store this information for a long (or short) period, and for what purpose. Within an ongoing project we report the result of searching and archiving activities of 30 academic researchers, mainly from biology and chemistry fields. We used contextual interviews, 10-day diary keeping and job-shadowing methods to understand researchers' current practices [5]. We observed in this study that amongst

most researchers there was a commonality in the patterns and goals of searching. However, goals and patterns of archiving differed to a greater extent. Based on the results we provide design implications for supporting efficient searching and archiving activities. In section 2 and 3, respectively, we will discuss goals and patterns of searching and archiving activities. In Section 4, we will discuss some implication for designing better searching and archiving support for digital technologies.

2 Searching

From the user studies it was clear that *goals* of searching online peer-reviewed material had a lot of similarities amongst the subjects. Researchers needed to search for relevant papers and other information to support their professional work activities like paper writing, teaching and supervising students, and collaborating with peers. As one of the researchers said, "*always… the goal is to find material that helps in writing an article*." On many occasions searching activities were triggered by the need to look for very specific information or data. Especially, researchers from Biology and Chemistry background needed to look for chemical structures and properties, especially their spectrograms, molecular weight, and other specific information. Many researchers believed that getting a quick look at the chemical reaction information of certain chemical structures could be "*very good to see exactly what they are doing in the paper without having to go through it*". Interestingly, researchers also wanted to get partial information during the searching activities. Easily extractable data like graphs, tables, and images were used for creating presentations slides for teaching. Sometimes the goal of searching was merely situation dependant. As one researcher reported, "*depends… Recently I am writing, so I quite frequently search, several times per week. When I write up the experimental section of a paper I don't need to search.*" Being at the top of the new literature was also an important goal for searching amongst the senior researchers who frequently write research proposals in order to get funding.

The *patterns* of searching document-based information differed amongst a few researchers. These patterns were mainly in the form of depth-first and breadth-first search. Some researcher spent quite a long time formulating the query before starting the search, whereas many others preferred a quick search and then refined it afterwards. Researchers mainly used publication title, author names, year, and keywords for searching. Interestingly, most of them reported the issue of having to use more than one tools to search for the required information. E.g., one pattern of using multiple sources to search was described as: "*I use SciFinder for searching chemical structures, CrossFire Beilstein to get chemical properties and the library system to retrieve full text.*" Some preferred using abstracting databases (instead of using full-text search data-bases like ScienceDirect) as these have more coverage and they could use more specific keywords and consequently get more relevant results. Some researchers preferred reading abstracts of papers in order to decide the relevance of the paper: "*I use the abstract to see whether it is a theoretical, empirical or normative paper.*" Interestingly, on many occasions the act of searching was triggered either by colleagues or email alerts. Several researchers had registered for

alerts on different online databases using certain keywords, so whenever a new document is published within their topics of interest the email can link researchers to the new document.

3 Archiving

Unlike the searching activity, the goals and patterns of the archiving activity differed considerably between different researchers. A major *goal* for archiving document-based information was to support the professional work. Most of the time online and paper-based documents were archived for the purpose of writing new papers and maintaining reference lists. Many of the web-based tools offer support for adding references to reference manager tools such as EndNotes and BibTeX. In other cases documents were stored for utilizing their partial information like images, graphs, charts, tables and other objects to use them in PowerPoint presentations for teaching and seminar purposes. Goals of archiving also included to support remote (away from work or while traveling) and rapid (printed papers for a quick reading) access. However, the goal of archiving was not limited to storing and retrieving information. As a secondary goal, archiving activities were also used for sharing and collaboration purposes. Some senior researchers collected and bookmarked papers to help their students. There were some personal issues behind creating a digital or paper-based archive. Although most of the material could be found online, some researchers wanted to archive them to have a backup. To avoid loosing valuable information some researchers tend to backup their data on a CD or a separate disk. One said, "*Very often, I tend to lose things… Another reason for me is when I am dialing in and suddenly lose my connection, and then I have to go back to the article again…*" Secondly, we observed that an archiving process was used as a mechanism to build identity within research groups. In one case a senior researcher, who was also the head of a research group, kept binders full of papers titled " experiment protocols" , "Plasmids Protocols", "Bench 2001-2002", "Sequences" and so on in a laboratory. He also kept an experimental notebook with data that only he would be able to decipher.

The *patterns* of archiving were mainly seen either in digital or physical forms, though we also found a mix of both. Documents were mainly stored in windows folders, email folders (alerts) and attachments, bookmarks, physical folders and binders on the desks or in the office wardrobes or sometimes just piled at different places in the office. Some people kept very important references in their backpacks and on USB sticks. In case of digital archives, most researchers archived document information in reference managers like EndNote or BibTeX. Researchers tend to download most documents they find relevant or interesting to their work. However, organizing these documents was a problem in most cases. Some researchers mentioned that their reference manager were not up-to-date. One researcher said, "*On disk I have twice as many articles as listed in BibTeX*." To organize their digital archives, researchers created folders on their disks for different topics (e.g. "cancer", "chemistry", "DNA targeting agent"). "*I have a directory 'projects', in there 6-7 separate projects by name (new paper, data files), in there manuscripts or correspondence, tables, figures, etc. When a project is finished I move it to a dead directory. I name all files clearly so I can find them by searching on filename*

(e.g. name of collaborator)." With a specific pattern of archive, one researcher said, "*I now simply use a unique code that indicates the topic, combined with a number. E.g. C001 is the first paper about Cancer. I use these codes in EndNote as well, and these are also used in ProCite in the lab. Then whenever I search in EndNote on some keyword, and I find a paper, I also have its number and I can retrieve the paper copy*." Some researchers tried organizing papers by structured naming schemes consisting of <title>, <author>, <year> and so on. Even when naming schemes were not always consistent, researchers thought they could easily find the documents by searching the file names. Some patterns of archiving partial information were also seen. Tools such as ScienceDirect allow access to manuscript figures, graphs and tables. Researchers tend to save these objects in PowerPoint files and later used in PPT files or transparencies during presentations. Sometimes screen captures of papers (via Print Screen) were also stored in PowerPoint.

Many researchers preferred printing and physically storing documents because this allowed them to annotate and underline texts, supporting the claims of [4]. Additionally, physical archives served as reminders, considering its placements, tangibility and visibility [3]. "*I can put it away and pick it up later from the same position. Print reads better than online, when reading from the screen I overlook some words or typos. Print is easier to form a fast judgment of a paper*." Another researcher pointed out, "*I never go back online to retrieve an article again. I keep it myself, otherwise I forget about it or I may not be able to find it again*."

It was observed that some personal and situational issues also influenced the way archives were created. Some senior researchers preferred archiving the printed versions rather than digital versions mainly to avoid weakening their eyes. Some didn't have access to the computer or network during travel so they preferred having printed versions in advance. One researcher wanted to minimize the paper load on her desk hence kept an archive of only the abstracts.

In some cases we observed a mix of physical and digital archives. A few researchers used note-cards to keep record of the important references stored in the digital form. The note-card writing allowed them to quickly make a record of relevant papers and to quickly draw relevant chemical structures. Although many of the subjects reported that they were not so happy with the way they archived their documents, they indirectly pointed several benefits of their archiving activities. For example, the printed archives allowed subjects to annotate and underline important issues which eventually allowed them to generate new ideas for their research.

4 Discussion and Design Implications

In the literature on PIM, it is always assumed that the material is already available either in physical or digital forms. Understanding *why* and *how* researchers search and archive document-based information resulted in a rich set of goals and patterns of searching and archiving activities. Our study showed that supporting professional work (e.g. paper writing, teaching), collecting references and using partial information were the common goals of both searching and archiving activities. Moreover, it also gave us indications about the relationship of searching and archiving processes and researchers' overall work practices. Our ethnographic exploration

showed that the goals and patterns of searching and archiving processes were highly dependant on academic researchers' everyday work activities. For example, when researchers were writing articles most of them preferred having a printed version of documents so that they could make annotations and/or scribble diagrams to appropriate their research ideas. In normal situations they mainly stored their findings in the 'To Read' and 'My Documents' folder as a temporary storage. Additionally, researchers who travel often preferred printed versions over the digital versions.

In the following we provide some rough design implications to better support searching and archiving process in digital technology.

- *Extra Searching Support:* Considering our users from biology and chemistry fields it was observed that there was a need to support their searching activities with more dedicated search entries. Researchers needed diagrammatic information about chemical reactions and structures to make better judgments about the overall suitability of documents. We suggest to develop enhancements in search engines by adding more relevant specialized search fields.
- *Archiving Support Beyond Storage:* While there is a wide range of tools available to search for documents, users always establish their own ways to archive information so that they could retrieve them easily. Successful archiving support should not be restricted to providing only functional support but should also allow researchers' to build their own (or group-) identity, and to share it with others, and to make the archive safer for future retrieval.
- *Supporting Partial Information*: Researchers wanted to search and archive partial information from documents. This partial information plays an important role in supporting work related activities like teaching, giving seminars, and student support. Additionally, researchers look for images and tables from documents to easily decide if the paper is relevant or suitable. As one researcher commented, "*I do article searching on Pubmed and search for methods and procedures, paying attention to images that show results*."

References

1. Barreau, D. and Nardi, B.: Finding and reminding: Organization of information from the desktop. SIGCHI Bulletin 27 (1995) 39–45.
2. Kaye, J., Vertesi, J., Avery, S., Dafoe, A., David, S., Onaga, L., Rosero, I., and Pinch, T.: To have and to Hold: Exploring the Personal Archive. Proc. of CHI'06. ACM Press: NY (2006) 275-284.
3. Kidd, A.: Marks are on the Knowledge Worker. Proc. of CHI'94. ACM Press: NY (1994) 186-191.
4. Sellen, A. and Harper, R.: The Myth of the Paperless Offices. MIT Press (2002).
5. Vyas, D., de Groot, S. and van der Veer, G.C.: Understanding the Academic Environments: Developing Personas from Field-Studies. Proc. of ECCE'13, ACM Press: NY (2006) 119-120.
6. Whittaker, S. and Hirschberg, J.: The character, value, and management of personal paper archives. ACM TOCHI. 8, 2 (June), ACM Press: NY (2001) 150-170.