What do People Recall about their Documents? Implications for Desktop Search Tools

Tristan Blanc-Brude
INRIA Rocquencourt
78153 Le Chesnay cedex, France
Tel:33-6-62-71-35-85
E-mail: tristan.blancbrude@gmail.com

Dominique L. Scapin
INRIA Rocquencourt
78153 Le Chesnay cedex, France
Tel:33-1-39-63-58-92
E-mail: dominique.scapin@inria.fr

ABSTRACT
This study aims at finding out which attributes people actually recall about their own documents (electronic and paper), and what are the characteristics of their recall, in order to provide recommendations on how to improve tools allowing users to retrieve their electronic files more effectively and more easily. An experiment was conducted with fourteen participants at their workplace. They were asked first to recall features about one (or several) of their own work documents, and secondly to retrieve these documents. The difficulties encountered by the participants in retrieving their electronic documents support the need for better retrieval tools. More specifically, results of the recall task indicate which attributes are candidates for facilitating file retrieval and how search tools should use these attributes.

ACM Classification: H5.2 [Information interfaces and presentation]: User Interfaces. – Ergonomics.

General terms: Experimentation, Human Factors, Design

Keywords: Personal information retrieval, desktop search tools, user studies, human memory

INTRODUCTION
The amount of information that we store in our computers through our work activities and in our private life is considerable. With that ever-increasing amount, a significant problem is the management of that information, particularly its retrieval. Studies have pointed out the limitations of traditional systems based on the desktop metaphor and the folder hierarchy [14, 20]. Thus there is an obvious need for new tools that would allow users to manage their data more easily. The term *Personal Information Management* (*PIM*) refers to the research field addressing the way people manage their physical documents (books, notebooks, sheets, etc.) as well as their electronic documents (files, emails, Web pages, etc.) with the aim of designing tools that support the management of electronic documents (*PIM tools*). According to Boardman [4], the activity of PIM itself can be broken up into 4 sub-activities: acquisition, organization, maintenance and retrieval of personal information. In our study, we focused essentially on the retrieval sub-activity which constitutes one of the major issues of electronic document management. Indeed, because of their constant accumulation, our documents are often difficult to retrieve in the bulk of our hard disks. Moreover, certain authors (Lansdale [17]; Barreau and Nardi [2]; Jones et al. [15]; Bellotti et al. [3]) stress the fact that the main goal in the activity of document management is precisely to allow their later retrieval. The results obtained by Dumais et al. [8] also lead to that conclusion as they observe that users feel less the need to maintain a complex hierarchy of their documents when they have a powerful search tool allowing them to find their documents more easily (see also [5] for an extended discussion on the implications of their results).

In this paper, we present the specific issues tackled in our study, followed by the method used. After the description of the results, we focus on their design implications. The conclusion summarizes the study, as well as its limits and future research avenues.

THE ROLE OF DOCUMENT ATTRIBUTES IN THEIR RETRIEVAL
PIM tools, whether they are research prototypes or commercial systems, exploit various *attributes* of documents to allow, in particular, their retrieval by the user. By *attributes* we mean the various dimensions characterizing a document like its size, its title, its color etc.

For example, in the personal file systems of current operating systems, the retrieval is mainly based on two attributes: the location of the documents in the folder hierarchy and the name of the documents. Other PIM research prototypes that have been designed as alternatives to traditional tools, chose to exploit in priority other types of attributes like, for example, the time stamp of the documents (e.g., *Lifestreams*, [11]) or the project context of the documents

IUI'07, January 28–31, 2007, Honolulu, Hawaii, USA.

(e.g., *UMEA*, [16]). Also, particular PIM tools (e.g., *Presto*, [7]; *Haystack*, [1]) as well as built-in desktop operating systems search tools (e.g., Spotlight for Mac OS X Tiger; MS Windows Vista's search tool) try to make available to the user as many attributes as possible (e.g., file name, keywords, time, size etc).

Boardman [4], however, underlines the lack of evaluation and empirical or theoretical grounding in the design of the majority of PIM research prototypes. In fact these prototypes often rely on "radical invention", a method which has been criticized for preventing from the setting of a common research grounding [25]. As for built-in system search tools, studies show that most users only use them as a last resort, preferring to browse manually through the hierarchy [2, 21]. We argue that one of the main reasons for their lack of use is their insufficient usefulness and usability. Even with the new generation of search tools (e.g., Mac OS Spotlight, Google Desktop etc.) which are improved, especially in terms of response time, usability studies are needed.

In fact, particularly with regard to usefulness, the reasons why certain attributes are selected rather than others, and the way they are exploited by these various types of tools are not supported by empirical data. Moreover, in certain cases, the attributes selected correspond to technical file metadata that systems can extract easily but which do not necessarily make sense for non-expert users. For example, the attribute of size in bytes is often used by search tools. If this attribute is easy to obtain and exploit from a technological point of view, it does not form part of information relating to documents to which an average user pays the most attention. Consequently, one can hardly expect that users will memorize it and recall it when they attempt to retrieve their documents. We agree on this point with Ravasio et al. [21] for whom the reason why built-in search tools are not used could be that *"existing search tools have never been designed for average users but rather for experts"* (p. 173). We agree also with their conclusion: "*Technical file metadata is, for the most part, useless to users and should be replaced by more user-friendly attributes"* (p. 176).

There is thus a need to create a basic knowledge on users' needs and behaviors on which new search systems could be built. We think, like Lansdale [17], that the design of PIM tools, in order for them to be adapted to users, must be grounded on the investigation of the cognitive mechanisms underlying the activity of PIM. With regard to the sub-activity of retrieval, as users carry out their search on the basis of what they remember about their documents, the investigation of the cognitive activity of memory of documents attributes thus seems paramount.

Several experimental studies went in that direction by testing the capability of users to memorize and recall certain attributes (e.g., location, appearance, names etc) in the context of document retrieval tasks [6, 9, 14, 17, 18, 20]. If those studies led to very interesting results, they however present the disadvantage of a limited ecological validity of the material used. Indeed, undoubtedly because of the constraints peculiar to controlled studies, the experimental tasks did not involve documents belonging to the participants themselves, but documents collected by the experimenter, that participants had to familiarize with in the context of the experiment. Thus, these studies do not permit to answer the following question: what do users really remember about their "own" personal documents?

Gonçalves and Jorge [13] took a step in this direction. They asked users to tell the story of three of their own documents with the aim to identify which kind of attributes are the most often evoked and how these attributes relate to each others in the narration. However, their study does not report either on the reliability of the recall (i.e. what is the degree of accuracy of the recall?; does it depend on the type of attribute?), nor on the way the attributes are recalled (i.e. what precise information relating to an attribute the users remember, with what degree of accuracy is the recall expressed, what are the expressions used to express the recall?). We think that this knowledge is necessary to design retrieval tools being adapted to the memorization and recall capabilities of the users and exploiting them as much as possible. The aim of the study described here is to try answering these questions.

METHOD

Participants

The study was conducted with two groups of participants (seven researchers and seven members of the administrative staff of a research institute), at their workplace, in the context of the management of their own working documents (digital and physical ones).

Participants were selected in order to form two groups carrying out two qualitatively different types of work with the aim of obtaining a more representative sample and of testing the possible effects of the type of work activity on the memorization of working documents characteristics.

Procedure

Semi-directed interviews followed by an experiment were conducted during sessions (one hour and a half to two hours) in the office of the participants.

Participants were first interviewed about their work, their environment, and about how they manage their files and documents (both physical and electronic). That large amount of collected data is not described here, as it is not the focal point of the study. In the context of this study, the aim of this semi-directed interview was only to select target documents which were the focus of the experiment conducted next.

The experiment, focusing on one, two or three user-specific documents (paper and/or digital) was conducted in two phases: a "recall" phase followed by a "retrieval" phase.

Recall Phase. This main phase aimed at exploring the participants' memory of the documents attributes. This phase consisted of two sessions: a "free-recall" session in which participants were asked to express the main features that

they recalled about these documents, and a "cued-recall" session, in which, attribute by attribute, participants were asked whether or not they recalled individual attributes that were proposed to them, from a pre-established list of 11 attributes. This list was made up starting from the major attributes identified in the *PIM* literature (e.g., [8, 13, 18]) and completed by attributes considered relevant for users and thus likely to be memorized. These attributes were:

- Location: name and path of the electronic directory or physical location in the desk,

- Type or Format: file type (e.g., Word document) or format of paper (e.g., stapled A4 sheets),

- File Name: name given to the electronic file,

- Title: title within the document if one had been given,

- Size: in terms of number of pages, number of lines etc. not in bytes.

- Time: time of last usage of the document,

- Keywords: meaningful words within the document,

- Visual elements: existence of graphics, tables, colors within the document,

- Associated events: significant events that occurred in association with the last usage of the document (e.g., emails, telephone calls etc.),

- Links: documents that are related to the target document (e.g., previous versions, documents that participate to the same task etc.),

- Actions: operations performed on the document by the user (e.g., printing, inserting. etc.).

Retrieval Phase. In this phase, the participants were asked to actually find the documents in their own environment, with their current tools. This phase aimed at observing how easily and by which means (i.e. use of a search tool or browse through a hierarchy of directories) participants managed to retrieve their documents with their own system. This phase also aimed at enabling the experimenter to study the retrieved document in order to check the correctness of the elements recalled in the previous phase. It should be pointed out that the analysis of the search activity itself was not the focus of the study which avoided the methodological flaws inherent to that kind of approach [10, 17]. Consequently, the choice was not to monitor and analyze precisely the "retrieval" phase (e.g., the search process was not video captured).

Material

The selection of the documents was directed by the experimenter on the basis of information collected from the semi-directed interviews during which participants were asked which were the main tasks they carried out at work. Each selected document corresponded to a document which had been used during an occurrence of one of the evoked tasks (e.g., presentation slides used by a researcher for one of the conference talk he gave). Documents were selected in a collaborative process between the experimenter and the participant. First, the participant described the workflow of one his/her task, mentioning the different documents that were used for the task. Then the experimenter decided which specific document would be a target document for the experiment. The requirement for the experimenter in this process was to lead to a final collection of documents differing from each others along the following dimensions: type of material (paper or electronic), recency/frequency of use (old, recent or recurring) and depth of processing (created by the participant, simply modified or simply consulted).

Out of all 30 selected target documents, there were 9 paper documents and 21 electronic documents; 12 documents out of 30 were created by the user himself, 13 documents had only been consulted by the user and 5 had been consulted and modified; 16 were so called "old" documents (i.e. they had been created, consulted or modified more than 6 months ago); 5 were "recent" documents (they had been created, consulted or modified less than 6 months ago) and 9 were "recurring" documents (which were consulted regularly).

The various types of selected documents were:

- *Electronic:* 8 text documents (2 PDF, 6 Word), 5 Web pages (or set of Web pages), 4 presentations (3 PowerPoint, 1 SVG), 2 e-mails, 1 table (Excel) and 1 set of files contained in a Windows directory (i.e. the source code of an application).

- *Paper:* 4 sets of sheets (stapled or not), 4 books and 1 set of pages in a notebook.

Table 1 shows the distribution of the selected documents within the different dimensions we chose to look at.

		Old	***Recent***	***Recurring***
Consulted	***Electronic***		**- 1 .doc** **- 1 .html**	**- 1 .doc** **- 1 .html** **- 1 .pdf**
	Paper	**- 2 books** **- 2 sets of sheets**	**- 1 set of sheets**	**- 2 books** **- 1 sets of sheets**
Modified	***Electronic***	**- 1 .html**	**- 1 .doc**	**- 2 .html** **- 1 .xls**
Created	***Electronic***	**- 3 .ppt** **- 2 e-mails** **- 2 .doc** **- 1 .pdf** **- 1 .svg** **- 1 source code**	**- 1 .doc**	
	Paper	**- 1 set of pages**		

Table 1: Distribution of all the selected documents according to their recency/frequency of use, the depth of processing, and the type of material.

RESULTS

The examination of the data was mainly based on quantitative and qualitative analysis of answers to the recall phase and behaviors observed during the retrieval phase. The quantitative analyzes relate to: frequency of freely expressed documents attributes and cued documents attributes; the robustness of the recalled attributes (verified by comparing the expressed attributes versus the attributes of the actual documents); the performance during the retrieval task. The qualitative analyzes relate to: the exact nature of information relating to an attribute evoked by the participants; the means of expression used by the participants during the recall of the attributes; the strategies used by the participants to retrieve their documents (including comparisons between the previously recalled attributes and the attributes that were used for the actual retrieval of documents).

Recall Phase

Frequency of Recalled Attributes. Concerning the "free-recall", the main characteristics expressed freely by the participants were: (1) characteristics of the documents' textual content (e.g., abstract, structure, distinctive portions of text like the title etc.): in 71.4 % of all document descriptions; (2) visual elements (e.g., existence of graphics, pictures, colors etc.): 25%; (3) file type or document format (e.g., "table Excel", "book format A5" etc.): 21.4%.

The attributes that were recalled more marginally were the size of the document (i.e. number of pages), a link (i.e. a document that is linked to the target document) and the author of the document (in cases where a document was not created by the participant himself).

The average number of different attributes recalled by participant was only of 1.46 attributes.

Concerning the "cued-recall", the results show the types of attributes that were most frequently recalled from the cued list (i.e. the attributes that the participants could instantiate for a particular document and tried to recall whether or not their recall was exact).

Five attributes (among the eleven that were suggested to the participants) were recalled in every case (for all participants and documents). These attributes are: location, file type or document format, time of last usage, keywords and associated events. The others attributes were not always recalled (although they were part of the actual documents attributes).

Results (see Fig.1) also show that all attributes of the cued list have a high percentage of recall (they concern at least 70% of the documents). This suggests that these attributes are good descriptors of the documents from a user's point of view.

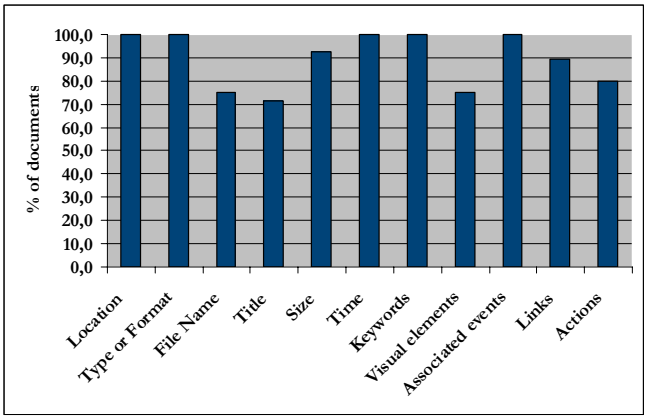

Figure 1: Percentages of documents for which attributes have been recalled.

The frequency of recall of attribute values on a document was related to the recency or frequency of use of that document only in the case of the attribute "visual elements". Indeed, "recent" and "recurring" documents all resulted in the recall (whether correct or not) of visual elements, which was not the case of the "old" documents. The calculation of the chi-square on the variables "type of document (old vs. recent/recurring)" and "frequency of the recall" indicates that these two variables are dependant ($X^2 = 6.03$, $p < .005$).

The results presented in the next sections relate to data concerning both sessions (i.e. the free-recall and the cued recall ones).

Correctness of Recalled Attributes. When possible (i.e. when document retrieval was successful), verification was performed by comparing the expressed attributes and the actual attributes of the document(s). The document attributes that were expressed but also (in most cases) characterized in the actual document were: the attribute "file type or format of paper" and the attribute "visual elements". Most other attributes are less robust or partly robust. For example, results indicate that only 36 % of the documents led to the recall of their exact location, although users currently manage all their documents this way.

Table 2 synthesizes the results concerning the correctness of the recall for each attribute except for: associated events, links and actions. It was very difficult indeed to check those because they relate to interactions between the document and its environment (the user, tasks, other documents etc.) which rarely left records in the system.

	Correct Recall	*Partial Recall*	*False Recall*
Location	**36 %**	**60 %**	**4 %**
Type or Format	**93.3 %**	**3.3 %**	**3.3 %**
File Name	**25 %**	**66.7 %**	**8.3%**
Title	**33.3 %**	**46.7 %**	**20 %**
Size	**15.4 %**	**30.8 %**	**53.8 %**
Time	**4.8 %**	**47.6 %**	**47.6 %**
Keywords	**32 %**	**68 %**	**0 %**
Visual Elements	**76.2 %**	**23.8 %**	**0 %**

Table 2: Percentages of documents for which the recall was correct, partial or false.

It is noticeable that a very important percentage of recall is in fact partial (e.g., keywords, name, location), i.e. only a part of the information recalled by the participants is correct but the other is not (e.g., often, for the same document, some recalled keywords were indeed part of the document while others were not), or the information is not accurate (e.g., a participant recalled the period corresponding to the last time he used the document but not the exact day).

Another result is that the least robust recalls relate to the attributes size and date.

The correctness of recall was related to the recency/frequency of use of the document only in the case of the attributes size and keywords. Concerning size, the proportion of correct recalls is much more important for recurring documents than for occasional (recent or old) documents. The calculation of the chi-square on the two variables "type of document" and "correctness of the recall" indicates that these two variables are dependant ($X^2 = 7.94$, $p < .001$), only when contrasting recurring vs. recent or old documents and correct vs. partial or false documents. With regard to keywords, the proportion of false recalled keywords is more important for old documents than for recurring documents. However the chi-square testing the dependence between the two variables is not significant even if it is very close to the threshold ($X^2 = 3.68$, NS or $p < .10$).

The correctness in recall of time was related to the type of participant but not significantly. False recalls are more frequent with the administrative staff than with researchers whereas correct recalls are more frequent with researchers. The chi-square tested on the two variables "type of participant" and "correctness of the recall" returns a value that is non significant but very close to the threshold ($X^2 = 3.83$, NS or $p < .10$) when contrasting correct or partial vs. false documents. This result could be explained by the fact that documents managed by the administrative staff are more often related to recurring routine procedures that can be more difficult to date compared with procedures related to particular events which are more frequent in researchers' work (e.g., preparation of a paper for a conference that took place on such date, slides used for a course given on such moment of the year, etc.).

Characteristics of Partial Recalls. Often a part of the recalled keywords, title or file name was correct but not entirely. Concerning keywords, all documents containing text led to the recall of, at least, one keyword that was indeed part of the document. But 68 % (17 out of 25) of the recall also contained keywords that were not in the document. The mean percentage of correct keywords compared with all recalled keywords (including cases where all keywords were correct) is 24 %; 32 % of the recalls (8 out of 25) only contained correct keywords. Concerning name, only 12.5 % of the partial recalls contained erroneous portions of text (versus 42.8 % for titles). The other partial recalls consisted in giving a subsection of the correct name.

The location recalled was often correct only to a certain extent, i.e. only a part of the path or location was correctly recalled. For electronic documents, the beginning of the path (the first directories) was correct, but not the end (last directories). In other words, the end of the path was either inaccurate (the name of the directories was partially recalled), plain wrong (wrong name or wrong folder) or was not recalled at all. In one case only the opposite occurred: one participant correctly recalled the name of the last folder but not the others. This concerned a directory located on a remote computer whose folder hierarchy had not been created by the user and whose path had to be typed with command lines. For paper documents, the partial recalls were those for which a location was provided (e.g., a cardboard, a book, a pile etc.) but the site inside this location was wrong or was not specified at all.

Size and time of last usage were often correctly approximated (e.g., in terms of intervals: "between 40 and 60 pages "; periods: "in February 2005" etc.) but rarely precisely (i.e. by recalling an exact number of pages or a precise day). For example, among documents for which the exact date of last usage could be checked, only one had led to the recall of the exact day.

Characteristics of False Recalls. Concerning the attributes of size and time of last usage, the distance between the recall and the verification was often proportional to the number of pages expressed or time elapsed. In other words, false recalls were not too far from reality.

For example, with regard to size, a participant said "around 50 pages" for a document that actually contained 55 pages. For this attribute, the average margin of error between the estimates and reality was 31.3 %. Furthermore, the difference between recall and reality is proportional to the size of the recalled number (the correlation is positive: $r(13) = .74$; $p < .004$).

With regard to time, the errors were made with a margin of error being of one day up to six months, but the majority was under two months. All the false recalls made during an estimate in terms of day were false with a margin of one day. The differences between the estimates in terms of month and reality were of two weeks up to two months. An error was made during an estimate in terms of year: it was inaccurate within 6 months. The difference between the recall and reality is thus proportional to the degree of accuracy of the estimate. The number of false recalls is not proportional to the time elapsed since the last usage because the more recent the document was, the more participants attempted to be precise.

Means of Expression Employed. A number of qualitative results were obtained about the characteristics and variability of the expressed attributes.

For example, qualitative results were obtained on the different means and types of variable units used by people to express approximations of size and time (e.g., "between 40 and 60 pages", "in February 2005" etc.).

With regard to size, in a little less than half of the cases (11 out of 26) participants have expressed their recall by giving an exact number whereas in the other cases (15 out of 26) they made an approximation. The approximations were of various types. In 46.6 % of cases, the participants gave a round number accompanied by a linguistic form marking the approximation (e.g., "approximately 40 slides", "about 500 pages"). The second most used approximation was the specification of an interval (e.g., "between 10 and 20 pages"): in 33.3 % of cases. Three other forms were used marginally: the expression of the lower limit (e.g., "at least 2000 pages"), the expression of the higher limit (e.g., "less than 100 lines") and finally by an adjective and/or an adverb (e.g., "a very short mail").

Concerning time, the dates were generally formulated in the form of periods (in 63.6 % of cases). Depending on the degree of precision, periods were expressed in terms of year, semester, season, month, week, day, half-day or hour. The periods were more often formulated in terms of month (e.g., "in February 2004"). The second manner of giving a date was an estimate of the time elapsed compared to present time in the form "…ago" (e.g., "two months ago", "about two weeks ago"): 30.3 % of cases.

With regard to the expression of file type, the participants did not formulate only the type from the point of view of the system but also from a semantic point of view of higher level (e.g., "a PowerPoint presentation", "an Excel table", "an image within Word", "a set of slides" etc).

Types of Recalled Information. Concerning visual elements, the various categories of recalled information were, from the most frequent to the least frequent: the color (e.g., "khaki green title", "burgundy red dots", "white background", "sky blue cover"); the format and layout (e.g., "double column", "title in big letters", "table of contents"); the presence of objects other than text (graphs, pictures, tables, etc.) and layout graphics (headings, strips, lines etc.). An interesting result is that the recalled elements did not only concern the first page of the document (only 10 % of all descriptions comprised only visual elements of the first page). 80 % of the descriptions contained both visual elements of the first page and visual elements of other pages. The remaining 10 % contained only visual elements of other pages. In particular, recalled objects (graphs, pictures, tables, etc.) were twice more often located "inside" the document than on the first page.

For actions, a frequent type of recalled information was the modification of the document. For a document created by a user, it means that it had undergone additional modification(s) after the creation of a stable version of the document (it does not concern the successive modifications inherent in the creation of any document). For a document not created by the user, that means that the user made its own modifications to the document (e.g., the filling of the fields of an electronic form, the writing of annotations on a paper document etc.). Three other types of frequent actions were: printing (for an electronic document); sending; and moving (i.e. changing the location of the document). The other types of actions recalled by the participants were, for the paper documents, photocopying, and for the electronic documents, uploading to the web, converting (i.e. from Word format to PDF) and inserting (of the document in another document).

With regard to associated events, 55.5 % of the recalled information can be linked to a software record directly related to the events. On the other hand, 44.5 % did not result in any software record. All records are in fact received or sent mails that are related to the purpose of the document and correspond, for the majority, to mail exchanged with colleagues. A little more than half of the records (mails) were still present on the machine of the participants while a little less than half were missing, in the majority of the cases because the participant had removed them. Recalled events that did not leave any record were mainly face-to-face or phone communications between the participant and colleagues. The other types of information were: the editing of a document and particular events (e.g., end of a project, etc.).

Concerning links, among all the recalled linked documents (N = 42), the majority (N = 37) was indeed present in the system for the electronic documents or in the office for the paper documents. The recalled documents were related to the target document, generally because they related to the same task or type of task. The majority of the target documents (20 out of 25) had at least a linked document whose link was not only "abstract" (i.e., in contents or function) but also "concrete" (i.e. in action or location) and thus potentially exploitable by a system for retrieval purposes (i.e. retrieving a document by following a link). An important number (35 % of the retrieved ones) of linked documents shared the same location as the target document: in the same directory for electronic documents and in the same folder or pile for paper documents. Another type of fre-

quent "concrete" link was a copy/paste operation (entering or outgoing). The other types of links were: hypertext link (entering or outgoing), version (previous or later), etc.

Retrieval Phase

With the methodology used, the results reported here for the retrieval phase only concern global behaviors, not detailed cognitive analyses, which were not the focus.

For electronic documents, the main method used for the retrieval was: browsing through the folder hierarchy, followed by a visual search within a directory. Only one participant used a text-based search engine, only because browsing-based search had failed. For paper documents, the strategy was to locate a particular space in the office, then to select the target document among other documents. In both electronic and physical worlds, the strategy is twofold: memory (recall) based search for a location, and recognition of the document among others. This distinction is coherent with the statement of Lansdale [17] that each information retrieval attempt implies two distinct psychological processes: a *recall-directed search* followed by a *recognition-based scanning*.

Five users did not succeed in finding a document (it concerned 4 paper documents and 1 electronic document). Out of the 25 documents actually found, 8 were not found easily including 2 paper documents and 6 electronic ones. In most cases (5 out of 8) the difficulty only concerned the *recall-directed search*: the user sought in the wrong directory or in the wrong location. In one case, it only concerned the *recognition-based scanning*: the target document was at the expected place, but it had to be recognized among many candidates (e.g., finding a mail in a folder). In the remaining two cases, participants had difficulties during both search phases.

DISCUSSION

Results show that participants often had difficulties in finding their own document(s) in their ordinary environment, with currently available tools. We observed that it is the first search phase that seems to be the cause of most problems. This phase being mainly based on the cognitive process of recall, the results collected concerning this process and their implications for design may help users in finding their documents more easily.

Results also show that when participants were asked to find their document(s), they only used a small subset of the attributes they were able to express (in the "cued-recall" session). Concerning the *recall-directed search* phase, a strategy of search based on location largely dominates. Regarding the *recognition-based scanning*, our methodology did not enable us to precisely analyze all the characteristics of the documents concerned in the process of recognition. Let us note however that this process seems mainly based, for paper documents, on the appearance and, for the electronic documents, on the recognition of the name.

Results of the retrieval task also confirm that built-in systems search tools are very rarely used. This supports the need for better tools for personal document retrieval.

When participants are asked to tell what they remembered about their documents, few items were recalled spontaneously; the majority of the recalled attributes were later induced by the experimenter, which restricts our results to the attributes that were considered to be important. This result is probably due to the well known fact that people hold more information in memory than they are able to recall: some information is *available* but not *accessible* [23, 24]. The presentation of the category of an item is likely to increase the recall of that item [12, 22]. It would be interesting to find a methodology that would elicit a greater proportion of spontaneous recall. For example, Gonçalves and Jorge [13] asked users to "tell the story" of their documents. This prompted the spontaneous recall of more attributes (a total of 17 attributes across all stories) than open question. However their goals were different and the methodology did not elicit detailed recalls for each attribute.

The fact that people recalled few attributes spontaneously but were able to recall them when prompted by the cued list indicates that the interface of search tools should explicitly suggest users to use the attributes that are likely to help document search and retrieval.

The "purpose" of a document, which can be considered as an important document attribute, is not part of our results for methodological reasons: the selection of the documents by the experimenter was made during the semi-directed interview, starting from the evocation of the participants' tasks; the purpose of the documents within tasks was thus recalled in an implicit way during this phase, and so could not be the subject of an objective later recall. In addition, the attribute "author" of a document could also have been added to our "cued-recall" list. Indeed, information relating to an author has been evoked for a document during "free-recall" session and the attribute was most of the time incidentally recalled by the user during the "cued-recall" session or the retrieval phase. Similarly, the people associated with the documents could also have been added to our cued list since they have been shown to be an important retrieval cue in personal information search [8, 5] and have often been recalled incidentally by the participants in our study.

Also, it should be pointed out that the way participants manage to recall attributes is not the only dimension to address when deciding how to use them in search tools. The relative capacity of attributes to discriminate documents from each others and to express queries that will return small sets of results is also an important dimension to address in order to design efficient search tools.

Despite these limits, results of the recall task provide information about the various attributes users are able to recall which, in turn, can lead to potential recommendations on how these attributes could be exploited by future retrieval systems.

It should be noted that suggesting recommendations does not mean that no existing tools implement some of them. The aim of such recommendations is rather to provide a basis for what it suitable and what is not, in order to contribute to the evaluation of existing tools and to guide the design of new ones.

These potential recommendations are presented below:

- Favor the best remembered attributes:

Attributes that are the most often and/or the most precisely recalled, namely location, file type or document format, time of last usage, keywords, associated events and visual elements, should be used in priority in retrieval tools.

Search by textual content: "keywords" is the only attribute which always led to a recall that contained at least one element (i.e. a word) indeed present in the document. That means that users have the capability to select, thanks to the recall of the textual content, a subset of their document in which the searched document will be present and to which they can apply filters to refine their search. This result goes in the direction of text-based desktop search engines (results also encourage to a lesser extent search by part of the file names since users often recall file names that are free from errors but incomplete).

Filter/Sort by type: the attribute type or format is, at the same time, one of the attributes that always led to a recall, and the one that is the most robust (it has the higher percentage of correct recalls). Therefore, it should be in first place in search tools filter/sort lists.

Recognize by appearance: results show that users almost always remember visual elements present in their documents and that they are very seldom mistaken on those characteristics. A preview of the appearance of the document itself, in addition to the simple presentation of the icon corresponding to the type of the document thus seems likely to facilitate the process of document recognition.

- Provide appropriate expressions of attributes:

Time expression: the most common ways used by participants to express time of last usage were: the specification of a period (a month in most of cases, e.g., "in February 2004") and the information on time elapsed (e.g., "two months ago"). Search tools that use time for documents retrieval should therefore allow these means of expression and provide them first.

Size expression: adverbs of approximation (e.g., "about 500 pages") and intervals (e.g., "between 10 and 20 pages") are the most common means of expression used by participants and therefore should be allowed by search systems and provided first.

Type expression: search tools should enable users to filter and sort documents by using higher level categories that make sense for them (e.g., a presentation, a table etc.) and not only those that are system based (e.g. .ppt, .xls etc.).

- Provide flexibility for attribute specification:

Wording flexibility: the fact that users are good at remembering keywords goes in the direction of the use of desktop search engines. But desktop search engines should be designed to accommodate the fact that users recall not only correct keywords but also erroneous ones for the same document. For example, the system could provide multiple text entry fields with independent enabling/disabling options for each one in order for the user to easily test new combinations of words when search results are not satisfactory; provide suggestions for synonyms; suggestions for super-ordinates; etc. Similarly, users often remember parts of file names and titles so systems must allow searching by parts (of names and titles).

Size flexibility: provide several types of variable size units (e.g., number of pages, number of lines etc.) and several means to express approximations (e.g., "between … and …pages"; "about … pages" etc.) to accommodate the variability of users expressions.

Time flexibility: provide several types of variable time units (e.g., year, month, weeks, etc.) and several means to express approximations (e.g., "used…ago", "during the month of…", etc....) to accommodate the variability of users expressions

- Provide extensibility for search results:

Location extensibility: given that users recall parent directories more easily than the subdirectories that actually contain the target document, provide visualization means to expand the view of directories to show files hidden in subdirectories.

Time extensibility: the system should provide ways of expanding the first estimation expressed by users given that their recall is often approximate. The extensibility should be proportional to the type of time unit expressed (e.g., if the user specifies a month period, system should allow visualizing also files that were used during the adjacent months). It should be noted that this margin of error appears all the more necessary for users who carry out routine tasks, which are not associated to memorable and easily datable events.

Size extensibility: given that users' recall is also often approximate, system should give the possibility to visualize also files that have more or less the same size. Our results suggest that an error tolerance of + or - 35 % would be suitable since the difference between the recalled value and the real value was, on average, equivalent to more or less than 37.8 % of the recalled value. In our case, the implementation of such a tolerance would recover more than half of the relevant document rejections on a size-based search.

- Provide visualizations of file contents:

With regard to document previews, the systems should not be limited to the first page because the recalled visual elements can just as easily relate to other pages of a document. Certain working papers can even be difficult to discriminate from each other by the first page because it is often standard (e.g., the first page of a scientific article). On the other hand, the objects specific to a document and likely to facilitate its recognition are often located in the other pages. Therefore, search systems should allow automatic or manual search of document elements that are included in a document (not only on the first page), for instance, provide "previews" of the document allowing for the search of a particular set of graphics, pictures etc., on all pages of the document.

- Provide explicit relationships between documents:

Given that users are able to recall documents associated with the document to be retrieved, systems should be able to record these associations so the user could retrieve a document starting from an (already found) associate one (e.g., enabling the retrieval of a file by showing copy/paste relationships between files; reduce the distance between files and emails management to make their relationships more exploitable for retrieval).

- Provide log of past operations on a document:

Given that users are able to recall actions that they performed on a document, other than opening or modifying, system should be able to record these actions and to enable their use, combined with other attributes, for file retrieval (e.g., search for a document printed in February).

- Provide semi-automatic combination of attributes:

The design of search tools could exploit the fact that the recall of certain attributes depends on the recency/frequency of use of the files. The system could foster certain association of attributes during search specification and discourage others. For example, if the user specifies a "date of last usage" that is quite old (e.g., more than six month), the system could suggest to enrich the search with those attributes which recall is not time-dependant (e.g., type, name, location etc.) as opposed to those which recall decreases with time elapsed (e.g., size, visual elements).

CONCLUSION

First of all, when participants were asked to find their own document(s), they only used a small subset of the attributes they were able to recall, and often they had difficulties in finding their document(s), with current available tools. This shows the need for better tools for retrieval. In addition, the results indicate which document attributes are more often recalled (e.g., keywords) but also which ones are best recalled, that is with less errors (e.g., type). Thus, results indicate which attributes are candidate for facilitating file retrieval. The results also show how the various attributes that the users are able to recall should be exploited to be usable. Systems should allow users to formulate the document attributes with the expressions and the degree of accuracy that characterize their recall. In addition, systems should take into account the approximate but foreseeable nature of the recall in the returned results by including a margin of error or allowing users to easily modify the parameters of the attributes. The results also provide ideas to exploit the fact that the recall of certain attributes can depend, on the one hand, on the type of user and, on the other hand, on the recency and/or frequency of use of the documents. Lastly, they encourage the tracking by the system of the attributes which relate to the interactions between the document and its environment (i.e., usage context events, links between documents and actions performed on the document) and they suggest which precise types of interactions are relevant to monitor and save.

Providing further analyses are conducted, together with additional controlled experiments, the recommendations provided in this paper can be viewed as a contribution to the design of improved retrieval tools. Obviously implementing these recommendations in actual PIM tools will need further development, ecological testing of their usability and utility, as well as comparisons with current systems, with a user-centered perspective.

ACKNOWLEDGMENTS

The authors would like to thank all the members of the INRIA staff who took part in this study.

REFERENCES

1. Adar, E., Kargar, D., and Stein. L. A. (1999). Haystack: per-user information environments. In *Proceedings of the eighth international conference on Information and knowledge management*, 413-422.
2. Barreau, D. and Nardi, B. (1995). Finding and Reminding: File Organization from the Desktop. *SIGCHI Bulletin*, 27(3), 39-43.
3. Bellotti, V., Ducheneaut, N., Howard, M., Smith, I., and Neuwirth, C. (2002). Innovation in extremis: evolving an application for the critical work of email and information management. In *Proceedings of the conference on Designing interactive systems*, 181-192.
4. Boardman, R. (2004). *Improving Tool Support for Personal Information Management.* PhD Dissertation. Imperial College, London.
5. Cutrell, E., Dumais, S. T., and Teevan, J. (2006). Searching to eliminate personal information management. *Communications of the ACM*, 49(1), 58-64.
6. Czerwinski, M., van Dantzich, M., Robertson, G. G., and Hoffman, H. (1999). The contribution of thumbnail image, mouse-over text and spatial location memory to web page retrieval in 3D. In *Proceedings of Interact '99,* 163-170.
7. Dourish, P., Edwards, W. K., LaMarca, A., and Salisbury M. (1999). Presto: an experimental architecture for fluid interactive document spaces. *ACM Transac-*

tions on Computer-Human Interaction (TOCHI), 6(2), 133-161.

8. Dumais, S. T., Cutrell, E., Cadiz, J. J., Jancke, G., Sarin, R., and Robbins, D. C. (2003). Stuff I've Seen: A system for personal information retrieval and re-use. In *Proceedings of SIGIR'03,* 72-79.

9. Dumais, S. T. and Jones W. P. (1985). A comparison of symbolic and spatial filing. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, 127-130.

10. Fertig, S., Freeman, E., and Gelernter, D. (1996). "Finding and reminding" reconsidered. *ACM SIGCHI Bulletin*, 28(1), 66–69.

11. Freeman, E. and Gelernter D. (1996). Lifestreams: a storage model for personal data. *ACM SIGMOD Record*, 25(1), 80-86.

12. Fox, P. W., Blick, K. A., and Bilodeau, E. A. (1964). Stimulation and prediction of verbal recall and misrecall. *Journal of experimental psychology,* 68, 321-322.

13. Gonçalves, D. and Jorge, J. A. (2004). Describing documents: what can users tell us? In *Proceedings of the 9th international conference on Intelligent User Interfaces,* 247-249.

14. Jones, W. P. and Dumais, S. T. (1986). The Spatial Metaphor for User Interfaces: Experimental Tests of Reference by Location versus Name. *ACM Transactions on Office Information Systems,* 4(1), 42-63.

15. Jones, W. P., Dumais, S., and Bruce H. (2002). Once found, what next?: A study of "keeping" behaviors in the personal use of web information. In *Proceedings of ASIST 2002,* 391-402.

16. Kaptelinin, V. (2003). Umea: Translating interaction histories into project contexts. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, 353-360.

17. Lansdale, M. W. (1988). The psychology of personal information management. *Applied Ergonomics*, 19(1), 55-66.

18. Lansdale, M. W. (1991). Remembering about documents: memory for appearance, format, and location. *Ergonomics,* 34(8), 1161-1178.

19. Lansdale, M. W., Simpson, M., and Stroud, T. R. M. (1990). A Comparison of Words and Icons as External Memory Aids in an Information Retrieval Task. In *Behaviour and Information Technology,* 9(2), 111-131.

20. Marsden, G. and Cairns, D. E. (2003). Improving the Usability of the Hierarchical File System, *Proceedings of SAICSIT 2003,* 122-129.

21. Ravasio, P., Schär, S. G., and Krueger, H. (2004). In pursuit of desktop evolution: User problems and practices with modern desktop systems. *ACM Transactions on Computer-Human Interaction,* 11(2), 156-180.

22. Tulving, E. (1983). *Elements of episodic memory*. Oxford : OUP.

23. Tulving, E., and Pearlstone, Z. (1966). Availability versus accessibility of information in memory for words. *Journal of Verbal Learning and Verbal Behavior, 5,* 381-391.

24. Tulving E., and Thomson, D. M. (1973). Encoding specificity and retrieval processes in episodic memory. *Psychological Review, 80,* 352-373.

25. Whittaker, S., Terveen, L., and Nardi, B. A. (2000). Let's stop pushing the envelope and start addressing it: a reference task agenda for HCI. *Human Computer Interaction*, 15(2-3), 75-106.