Exploring Memory in Email Re-finding

DAVID ELSWEILER, MARK BAILLIE and IAN RUTHVEN
Department of Computer and Information Sciences, University of Strathclyde

Human memory plays an important role in personal information management (PIM). Several scholars have noted that people re-find information based on what they remember and it has been shown that people adapt their management strategies to compensate for the limitations of memory. Nevertheless, little is known about what people tend to remember about their personal information and how they use their memories to re-find. The aim of this article is to increase our understanding of the role that memory plays in the process of re-finding personal information. Concentrating on email re-finding, we report on a user study that investigates what attributes of email messages participants remember when trying to re-find. We look at how the attributes change in different scenarios and examine the factors which impact on what is remembered.

Categories and Subject Descriptors: H.3 [**INFORMATION STORAGE AND RETRIEVAL**]: Personal Information Management

General Terms: Design, Human Factors

Additional Key Words and Phrases: email re-finding, information re-finding, memory, user study

1. INTRODUCTION

As the amount of digital information people create and use has increased, a need has developed for tools to help with the management of personal information rather than relying simply on memory alone. The tools currently available to help people re-find previously stored information are mostly either search-based, where the user can create queries based on attributes they can remember, or browse-based, where the user can manage their information objects by imposing their own organisation on the collection e.g. folders in an email client. However, despite these tools being created so that the user does not have to hold all of their information in their head, it has been suggested that with both the searching and managing approaches, the load for successful recovery of information remains on the user's memory [Elsweiler et al. 2007].

To conduct a successful search on a query-based system such as Google desktop, for example, a user must remember sufficient details about the information they want to retrieve in order to form a query. Studies of recollection for texts and stories, however, indicate that people are not good at remembering precise details. Instead what tends to be remembered are high-level meanings or gists [Sachs 1967; Clark and Clark 1977; Rubin 1977]. The findings of these studies suggest that

Author's address: D. Elsweiler, Department of Computer and Information Sciences, University of Strathclyde, Livingstone Tower, 26 Richmond Street, Glasgow G1 1XH

people would not be adept at remembering terms in a document, the subject of an email etc. - the kind of recollections required to construct queries.

The major alternatives to query-based systems are browse-based systems in which a user looks through information objects in order to find the objects they want. Browsing systems either show users all of the objects available, limiting the approach to relatively small data sets, or force a classification on the objects such as colour distribution for images [Heesch and Rüger 2004], concepts for documents [Yang 1994], etc. Similarly, information management tools force a classification on users, either by automatically classifying objects, as in text categorisation systems [Hayes et al. 1990], or by forcing users to classify objects, usually in some form of hierarchical system [Malone 1983]. For example, photographs and music are generally organised in albums and possibly further sub-categorised by artist, date, genre etc. Operating systems manage applications and files in a hierarchical system of folders, email tools provide facilities to group messages hierarchically, and standard web page book-marking features are hierarchical. Despite their popularity, hierarchical systems have been shown to have problems. Malone's study of natural office behaviour demonstrated that they are cognitively challenging and that users are reluctant to use them either because they cannot decide how to categorise an item, or because they are not confident in their ability to recollect at a later date how an item was categorised [Malone 1983]. Similar behaviour has been observed with digital documents [Boardman et al. 2003] and email messages [Whittaker and Sidner 1996]. Recently, systems that allow the user to attach semantic keywords or "tags" to documents have been proposed as a potential solution to the limitations of browsing systems [Cutrell et al. 2006], as have faceted-browsing systems [Yee et al. 2003; Schraefel et al. 2005]. We have also seen the development of hybrid systems that merge the interaction modes of browsing and searching. However, much has still to be learned about the benefits of these systems in a PIM context, how they are used in practice and, indeed, how they support the memory of the user.

The limitations of existing Personal Information Management (PIM) tools and growing quantities of personal information combine to motivate our work. In particular, we are interested in the role that human memory plays in the management of personal information. In PIM people try to obtain information based on the features of an object that they can recall [Capra and Perez-Quinones 2005]. Therefore, to improve the utility of PIM tools, we believe that what is needed is a greater understanding of what people can remember and how tools can be designed to better support these memories.

This article investigates the role that human memory plays in email re-finding. It reports on some of the findings of a user study that was designed to investigate what people remember when they re-find email messages and how different kinds of systems support these recollections. The focus of this article is on the recollection data, which are analysed in two steps. First, the data are examined to determine which attributes of email messages the participants remember and how these change in scenarios that should have an effect according to previous research. Second, the data are analysed statistically to discover which factors influenced the attributes that were remembered. By examining memory in this way we show what kinds of information people have available to help them search - the kinds of information

that re-finding tools should support.

The remainder of the article is structured as follows: Section 2 provides the relevant background literature; section 3 outlines the methodology employed and the technique used to examine what the participants remembered. Section 4 presents the recollection data. Section 5 examines these data statistically to determine the factors influencing what the participants remembered. In sections 6 and 7 we outline the limitations of the study and discuss the findings. Finally, our conclusions are presented in section 8 set against the context of future work.

2. RELATED WORK

This section describes the background literature for the primary themes of this article. Section 2.1 motivates the examination of memory in email re-finding. It introduces the problem of email management and summarises previous work in this area; section 2.2 presents work that has related memory and PIM; and section 2.3 presents findings from the field of cognitive psychology that warrant further investigation in a PIM context.

2.1 Email Management

Although email originated as an asynchronous communication tool that allowed text-based messages to be sent between users, recent research has confirmed that email is used for a variety of purposes. It has been shown that email is used for collaborative working [Ducheneaut and Bellotti 2001], data archiving [Mackay 1988; Whittaker and Sidner 1996; Bälter 2000], as well as the management of tasks [Whittaker and Sidner 1996] and personal contacts [Whittaker et al. 2002]. Therefore, rather than deleting email messages as they are received and read, users tend to keep their messages for future purposes and, consequently, many users have collections consisting of thousands of messages [Whittaker and Sidner 1996]. Studies have also revealed differences in the strategies used to organise personal email collections. For example, Mackay [1988] distinguished between "prioritizers" who organise their messages in such a way that they can give priority to messages that require attention and "archivers" who are mainly concerned with keeping information in case of future need. Whittaker and Sidner [1996] identified three user strategies based on the use of folders and the frequency with which the collection is maintained. Some users (*no- filers*) make no use of folders and instead leave all of their messages in the inbox; other users (*frequent filers*) use folders as a means of placing an organisation on the collection and make efforts to regularly sort messages in their inbox into appropriate folders; and a final group of users (*spring-cleaners*) also make use of folders, but only clean up their inbox periodically. The evidence suggests that there is a tension between the way people organise their messages for the different email activities described above. For example, when using email for task management purposes, a common strategy is to leave an email message in the inbox to act as a reminder to perform a particular task [Mackay 1988; Whittaker and Sidner 1996]. This strategy may be less successful for "no-filers" who will have lots of messages in the inbox, or may hinder re-finding attempts of "filers" who normally files messages into folders. All of these factors - large inboxes with thousands of messages, multiple uses of email, multiple and conflicting filing strategies - combine to place a huge burden on the user's memory when re-finding. This makes email a

particularly important media to study with respect to our research aims.

2.2 Memory and PIM

Previous research has shown the role of memory in PIM to be non-trivial. For example, Lansdale [1988] described office organisational problems as problems of categorisation, recognition and recollection; Case [1991] proposed that memory and metaphor impact the way historians manage their resources; and Carroll [1982] demonstrated that simple eight character file names can trigger a detailed recollection of a file's content. It has also been observed that memory problems and the limitations of human memory hinder PIM [Jones et al. 2005; Czerwinski and Horvitz 2002; Elsweiler et al. 2007].

Consequently, many groups have attempted to design systems that support known characteristics of memory. For example, the systems designed by Freeman and Gelernter [1996] and Ringel et al. [2003], amongst others, attempt to leverage episodic memories and the fact that events are framed temporally with respect to the times of other events; the systems designed by Kaptelinin [2003] and Jones et al. [2005] exploit strong human abilities to relate information objects to contexts in which they were created or used; the Placeless Documents system [Dourish et al. 2000] exploits the fact that attributes of documents may be remembered better than their storage location, and the systems designed by Dumais et al. [2003] and Cutrell et al. [2006] exploit the fact that people usually find it easier to recognise than to remember. All of these projects are credible attempts to leverage psychological research to improve PIM tools. However, each group interprets memory research slightly differently, focusing on specific findings. There is little evidence in the psychology literature that suggests utilising one type of memory over another in the context of re-finding information e.g. to exploit temporal memories rather than semantic memories or vice versa. There is, however, some evidence from PIM studies that reveal clues about memory in the context of PIM. For example, the finding that users prefer to locate their documents spatially rather than using keyword search [Barreau and Nardi 1995] suggests that spatial memory can be useful for re-finding. There is also evidence that spatial memory can be better exploited when the document space is three-dimensional [Robertson et al. 1998]. However, Jones and Dumais [1986] warn against over-reliance on a spatial organisation. Their findings indicate that semantic labels provide stronger retrieval cues than spatial organisation alone, although combinations of semantic and spatial organisation can enhance performance. Lansdale and Simpson [1990] extended this finding by discovering that semantic and spatial cues are enhanced when the user selects the cues themselves, rather than having them selected by an external party. There is also evidence in the PIM literature for the utility of temporal and episodic memories. Ringel et al. [2003] discovered that people can relate documents to events that happened around the time that the documents were created or used and that this can be utilised when re-finding. Further, examining the interaction log files of re-finding tools shows that users often remember that documents are connected to particular people and use these memories when creating re-finding queries [Dumais et al. 2003; Cutrell et al. 2006].

While these studies are useful and add to our knowledge of the role that memory plays in PIM, the findings relate to specific tools and do not report concretely on

what people actually remember. Therefore, they do not provide sufficient evidence to advocate supporting one particular feature of memory over another in PIM tools. We argue that in order to discover how best to design re-finding tools and to determine which attributes of memory to support, there is a need to understand the kinds of attributes people remember when they are trying to re-find irrespective of the tool that they are using. This is the approach taken by both Gonçalves and Jorge [2004] and Blanc-Brude and Scapin [2007]. Gonçalves and Jorge [2004] asked participants to tell stories about three of their personal documents by describing each, from memory, in terms of its features, its content and the context in which it was created or used. It was discovered that time, location, and purpose of the document were the most common attributes used in stories. Similarly, Blanc-Brude and Scapin [2007] used semi-structured interviews to examine participants' recollection of their documents. They found that location, format, time, keywords and associated events were remembered most frequently, but many of these attributes, particularly keywords, time and location were often only partially remembered or the recollections offered by the participants were incorrect. We believe that these studies provide a good starting point for investigating the role of memory in PIM, but what they provide is a very general snapshot of the attributes that the participants remembered. It seems likely to us that people will remember different types of documents in different ways because different types of document (e.g. email messages, photographs, spreadsheets, web pages, and word processing documents) are created or attained in different ways and are used for different purposes. Therefore, we decided to narrow the scope of our study by focusing on one type of document – email messages. Further, the psychology literature suggests that the attributes that people remember will not always be the same and will be influenced by a number of contextual factors. The following section provides a short review of relevant psychology literature that motivates investigating how particular factors affect what people remember about the information they are trying to find.

2.3 The factors that may influence memory in PIM

The user study presented in this article investigates the changes in what the participants remember depending on 4 factors:

—the time that had elapsed since the user last accessed the information
—the type of task being performed
—the type of user performing the task
—the filing strategy the user applies

Below, we outline previous research that motivates examining these factors.

2.3.1 *Time.* The transient nature of memory is well established. Ebbinghaus [1885] was perhaps the first to recognise that the quality of encoded memories degrade with time, but this has been confirmed by other scholars for different types of memory [Rubin and Wenzel 1996; Thompson et al. 1996]. There is also an abundance of discussion in the literature about episodic memory - a separate memory system that stores information as episodes. Episodes include information such as the location of an event, who was present, and what occurred before, during, and after the event [Tulving 1983]. Other research has demonstrated the temporal

encoding of information. For example, Brown et al. [1985] found that people are good at dating events, but subsequent research explained that these abilities rely on mnemonic strategies such as framing events in the context of other events, either historic or autobiographical [Huttenlocher and Prohaska 1997], which may or may not be possible within PIM. Therefore, although psychological evidence relating to temporal memories is relevant and potentially useful to the design of PIM systems, it is not entirely clear what these findings mean in the context of PIM. Before designing an email re-finding system based on the findings, it is important to discover if people actually do remember when emails were sent or indeed if what they remember changes as time goes by.

2.3.2 *Task.* In addition to a separate store for temporal memories, psychologists have proposed that different systems exist for episodic, semantic and procedural memories and these distinctions have been endorsed by neuroimaging analysis [Cohen 2004]. Other research indicates that memories can be organised in different ways, and information can be encoded visually [Kosslyn 1981], spatially [Kerr 1983], acoustically [Baddeley 1966], semantically [Grossman and Eagle 1970] and temporally [Brown et al. 1985]. The fact that memories can be created and stored in different ways suggests that it is plausible that information objects processed and used in different contexts will be remembered differently. Therefore, different attributes of email messages may be remembered in different situations and if this is true, various types of tasks, such as those proposed by Elsweiler and Ruthven [2007], may require particular kinds of tool support. Before tools are developed to assist with these tasks it would be useful to establish if different types of tasks are associated with specific remembered attributes.

2.3.3 *Type of User.* Psychology research also indicates that personalisation and self-reference can improve the quality of memories [Eysenck 1992; Rogers et al. 1977]. Further, previous experiences and knowledge have been shown to change the way memories are created and the later retrieved [D'Andrade 1995]. These findings suggest that different groups of users, each with different types of knowledge and personal experiences, may remember different attributes about their emails. If this is the case, different groups of users may benefit from different types of tools support, based on what they remember. Our study findings are analysed to determine if this is the case.

2.3.4 *The User's Filing Strategy.* The final factor to be investigated is how the email filing strategy applied by the user impacts on what is remembered about email messages. As noted above, Whittaker and Sidner [1996] observed three different email filing behaviours: no-filers, frequent filers and spring-cleaners.

The psychology literature suggests that employing a filing strategy should increase what people remember about their information. For example, people remember lists of words better when there is some kind of organisation imposed on the words [Bower et al. 1969], particularly when the organisation is devised by the person themselves [Wittrock and Carter 1975]. Further, according to the work of Craik and Lockhart [1972], the cognitive processing involved in filing would improve the user's memory of filed emails by increasing the quality of the encoding process. Nevertheless, another school of thought exists that suggests that the op-

posite effect would be true. Mander et al. [1992] proposed that people who do not file may need to interact with their information objects more often because they are forced to browse their collection in order to re-find. Mander suggests that this increased interaction with personal objects would facilitate an improved awareness of the information within a collection because it would mean interacting with emails that filers would not see again after filing. Consequently, by applying a non-filing strategy, Mander suggests that the user will actually remember more about the contents of their information collection. With respect to our aim of understanding the role of memory in PIM, this is an interesting argument. We examine the data collected to determine if there is any evidence supporting either case.

This section has presented the appropriate background literature for our work and motivated the study of memory in email re-finding. We have underlined the importance of human memory to PIM and proposed that a better understanding of what people remember may lead to the design of improved PIM tools. We have also presented findings from the field of psychology that we believe may impact on what people remember in a PIM context. The aim of this article is to investigate how these factors affect what people remember when re-finding email messages and interpret what this means in terms of how PIM tools should be designed. To achieve our aim we analysed the recollections participants had while taking part in a controlled user study of email re-finding behaviour. The following section outlines the methodology employed.

3. METHODOLOGY

We conducted a user study which examined the participants' recollections while they performed email re-finding tasks using three experimental systems: a browse-based system similar to the folder-based interface from Mozilla Thunderbird [1], a search-based interface similar to the search-based interface in the same Mozilla Thunderbird email client, and a third interface which was designed specifically to support memory based on previous investigatory work. The third interface offered the user a more visual form of interaction based on thumbnail images of the sender of messages and provided a means of interaction that was a mixture of browsing and searching. As the focus of this article is not on evaluating the performance of the systems used, but purely on the recollection data which were not influenced by the systems[2], we omit specific details on the interface designs.

The difficulties involved in performing controlled studies of re-finding behaviour are well documented [Boardman 2004; Capra and Perez-Quinones 2006; Elsweiler and Ruthven 2007; Kelly and Teevan 2007]. These include sourcing collections, creating experimental tasks, balancing the experimental design and protecting the privacy of users. The methodology we employed to overcome these challenges was that proposed by Elsweiler and Ruthven [2007], which allows the behaviour and performance of the participants to be evaluated while they complete **realistic** re-

[1] available from http://www.mozilla.com/en-US/thunderbird/

[2] Chi-squared tests for association show that the re-finding system used had no influence (statistical significance was considered at $p<0.05$) on the recollection of the four principal attributes: "when sent" (Chi-Sq=1.0, DF=2, p=0.61), "sender" (Chi-Sq=0.9, DF = 2, p=0.65), "topic" (Chi-Sq=5.0, DF = 2, p=0.08) and "reason sent" (Chi-Sq = 4.3, DF = 2, p=0.12)

finding tasks, on **real** (their own) collections, in the **controlled** environment of the laboratory. To generate personalised search tasks that can be performed on individual participants' collections, Elsweiler and Ruthven [2007] recommend that studies should be performed in two stages. The first stage involves conducting a diary-based study where a larger number of participants record details of the re-finding tasks that they are required to perform as they go about their normal daily activities. The tasks recorded and the details supplied about these tasks provide enough information that enables the experimenter to devise a set of experimental tasks that can be completed by the participants using their own personal collections. The final evaluation is then performed in a second stage with a subset of the original participants who volunteer to continue with their participation.

In the first phase of our empirical work, 34 participants, consisting of 16 post-graduate students and 18 undergraduate students, recorded details about 150 email tasks as they were completed over a period of approximately 3 weeks. Previous work has shown that people perform three kinds of email re-finding tasks: lookup tasks that involve finding specific information from within an email, item tasks where a single email is required, and multi-item tasks where information needs to be recovered from multiple emails [Elsweiler and Ruthven 2007]. Following the guidelines provided by Elsweiler and Ruthven [2007], we found patterns in the recorded tasks to learn about the activities that the participants used email for and established overlap between the participants. For example, participants in both groups recorded tasks relating to classes that they were taking at the time and often different participants recorded tasks that involved searching for the same information. This was useful because it provided us with a clue that even though some of the participants did not record a particular task, it was possible that the task may still be applicable to their collections. Other patterns revealed included the discovery that students within the same group often searched for emails containing announcements from the same source. For example, several undergraduate students recorded tasks that included re-finding information relating to job vacancies. There were also tasks that were recorded by participants in both groups. For example, searching for an email that would re-confirm the pin code required to access the computer labs.

To supplement our knowledge of the participants' email collections, we asked 2 participants from each group to provide email tours. These consisted of short 5-10 minute sessions, where participants were asked to explain why they use email, who sends them email, and their organisational strategies. This approach has been used successfully in the past as a non-intrusive means to learn about how people store and maintain their personal information [Kwasnik 1989]. Again, patterns emerged that helped with task creation. We found content overlap within and between groups that confirmed many of our observations from the diary study data. For example, the students who gave tours revealed that they received emails from lecturers for particular class assignments, receipts for completed assignments, and various announcements from systems support and about job vacancies. Importantly, the participants were also able to confirm which other students had received the same information. This confirmed that many of the tasks recorded during the diary study were applicable, not only to the recorder, but to every participant in one or both groups.

We used our acquired knowledge about how the participants used email to create a set of experimental tasks for each group of participants that included tasks of the three types outlined above. We also created a set of tasks for a third group of participants that consisted of research and academic staff members from the Department of Computer and Information Sciences at the University of Strathclyde who volunteered after receiving an email invitation to participate in the study. We created tasks for this third group of participants based on emails that were sent out to every staff member in our department. This task creation technique has been used successfully in the past to study email re-finding behaviour [Ringel et al. 2003]. There was no need to include participants from the researchers group in the earlier task analysis phase as we had previously learned about the tasks this group of participants perform in a previous study [Elsweiler and Ruthven 2007]. Where possible we used the information recorded in the diary study descriptions to provide a context for the tasks that we created i.e. a work task or motivation that would require the task to be performed. The task pools for each group of users can be found in Appendix A. The pools consisted of a mixture of tasks. Some tasks e.g. task B6 were applicable only to other participants in the same group, while others, such as task A1, were applicable to all three groups. Nevertheless, overall, the tasks in the three pools reflect the email usage patterns of three different user groups.

3.1 Participants

In the second phase, which started three weeks after the task recording diary study had completed, 21 participants were evaluated using their university email accounts. There were 7 participants from each of the three groups described above (undergraduate students, postgraduate students, and research staff). With the exception of the researcher group, each of the participants volunteered after completing the earlier task collection phase.

We structured the experiment around the user groups because of the differences between the email behaviours exhibited between the groups. The three groups of participants were very different. They had different numbers of email messages, used email for different purposes and had different levels of experience with using email. The properties of the participant groups are shown in Table I.

Property	Postgraduate	Undergraduate	Researchers
Number of emails (median)	106 (min=95, max=228)	187 (min=76, max=1165)	5132 (min=1097, max=8954)*
Age of oldest email (days)(mean)	76.15 (SD =2.44)	634 (SD=314.65)*	941 (SD=546.08)*
Number of filers	0	2	2
Number of No-filers	5	3	2
Number of spring-cleaners	2	2	3
Emails received per day (mean)	1.78 (SD = 0.70)	0.57 (SD =0.60)*	8.03 (SD = 4.51)*
Experience with using email	3 (IQR=1.25)	4 (IQR=0.25)*	4 (IQR=0.00) *

Table I. The email properties of the three participant groups. Statistically significant differences are indicated by *

The postgraduate group had not been enrolled at university for long and therefore had very few email messages. Similarly, the undergraduate group had low numbers of messages. This was because even though the undergraduates were recruited from the 3rd and 4th academic years and had been using their accounts for some time, the accounts had only recently been upgraded to the IMAP standard (where messages

are left on the server). The undergraduate participants did have on average more emails than the postgraduate students. However, the difference in collection sizes between the two groups was not significant. Participants in the researcher group, on the other hand, had significantly more emails than both the postgraduate and undergraduate participants.

Many of the postgraduate students came from non-computer science backgrounds and had low-levels of computer literacy, including limited experience with email. This was evident during the evaluation. From informal interviews it was discovered that the participants in this group generally used email for class announcements and reported that they had less need to re-find information. The undergraduate students were recruited from 3rd and 4th year classes and therefore, had much more experience with computers and using email. The main uses of their departmental email for the undergraduate participants were class announcements, university related task management, collaborative work, and social communication with university colleagues. Nearly all of the participants remarked that they also had other email accounts that they used for personal and non-university purposes. The participants from the researchers group were also very experienced email users, had been using their accounts for longer periods of time and therefore had large numbers of messages, used email as a way to manage their activities and their documents and as a result they reported having to re-find information often.

A pre-study questionnaire asked the participants to rate their experience with using email browsing and search facilities to re-find emails. The participants answered on a scale from 1 to 5 where 1 meant no experience, 2 meant limited experience, 3 meant average experience, 4 meant reasonably experienced and 5 meant very experienced. The questionnaire data show that the postgraduate participants had significantly less experience with email re-finding than the other groups. However, there was no evidence to suggest a difference in the experience levels between the participants in the undergraduate or the researcher groups.

The undergraduate participants received and kept the fewest emails of the groups. This was significantly fewer than the postgraduate group and the researcher group. Further investigation explained this by revealing that the undergraduate participants tended to exert more effort in collection maintenance than the other groups. The researchers received and kept the most emails of the three groups. However, all three groups in our study processed considerably lower quantities of messages than reported in other studies. For example, Fisher et al. [2006], Whittaker and Sidner [1996], and [Mackay 1988] all found that their participants received between 40 and 60 emails per day. The differences can be partially explained by the way the figures were calculated. Both Mackay [1988] and [Whittaker and Sidner 1996] asked participants to estimate the volumes of email they receive. We, on the other hand, calculated our figures based on the collections themselves by dividing the total number of messages by the total number of days passed since the date of the oldest email. Therefore our figures do not include any emails that were deleted, but do include holidays and weekends where the volume of email would likely have been much lower. Although we feel that the method of calculation accounts for some of the differences, it is fair to say that overall, our population had different characteristics to those of previous studies. We discuss this further in section 6.

To summarise, the three groups of participants in our study had very different characteristics. The postgraduates had low expertise, few messages and used email as a simple communication tool. The undergraduates had high expertise, few messages and used email for keeping track of class assignments and other university related tasks. Finally, the researchers had high expertise, lots of messages, but used their email for many purposes including managing content, scheduling, and task management.

3.2 Tasks

The tasks used in the evaluation were taken from the pools created in the first stage of the experiment. Each participant performed 9 tasks, with 3 tasks (1 lookup, 1 item, and 1 multi-item) being performed on each of the three experimental systems. The task types and systems were rotated to create a balanced experimental design [see Figure 1].

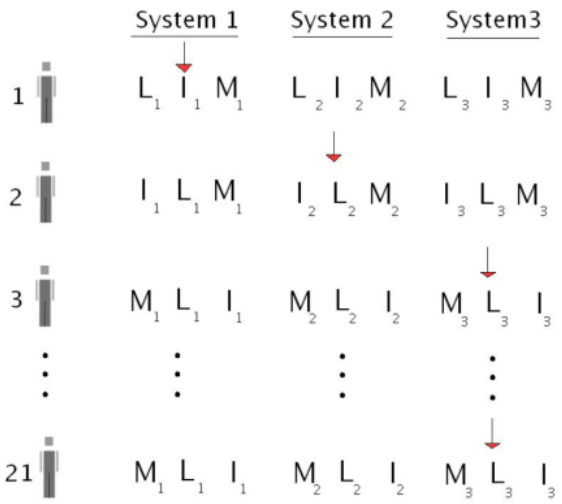

Fig. 1. A diagram showing how task types (lookup, item, multi-item) and systems were rotated. The arrows demonstrate the starting system for each participant

Before completing each task the participants answered questions about the task, including how clear they felt the task description was and how difficult they perceived the task to be. Both questions were answered on a scale from 1 to 5. The tasks were mostly rated as very clear (median = 5, IQR = 0 [3]) showing that the descriptions of tasks were understood, but there was a good mix of task difficulties (median = 2, IQR =1). Some of the tasks were perceived to be quite easy, while

[3]In descriptive statistics, the interquartile range (IQR) is the range between the third and first quartiles and is a measure of statistical dispersion. The interquartile range is a more stable statistic than the (total) range, and is often preferred to the latter statistic. A larger IQR means a larger range of data

others were considered challenging. Similar ratings were applied across the different user groups and filing groups.

To determine how the recollection data changed over time, before each task we asked participants to specify roughly how long it had been since they last accessed the information the task required them to re-find. Participants could choose between the following options: In the last day, in the last week, in the last month, in the last year, or over a year. To simplify the data analysis process and to align our work with previous PIM research, the scale was simplified to the temperature metaphor proposed by Sellen and Harper [2003]. Tasks that required information accessed in the last day or week were classified as hot, tasks that required information accessed in the last month were classified as warm and tasks requiring information last accessed over a month ago were classified as cold.

Of the 189 performed tasks, 45 were hot, 45 were warm and 60 were cold. The remainder of the tasks (mainly multi-item tasks) were classified as a temperature range i.e. different pieces of the sought-after information had been accessed more recently than others. Thus, although the way tasks were created and issued meant that we only had limited control over the temperature of the tasks, the issued tasks represented a reasonably balanced mix across the temperature range. The tasks were also balanced across the groups. See Table II below.

Temperature	Postgraduate	%	Undergraduate	%	Researcher	%
Hot	13	20.6	23	36.5	9	14.2
warm	26	41.3	12	19.1	7	11.1
Cold	8	12.7	25	39.7	27	42.9
Range	16	25.4	3	5.8	20	31.8
Total	63	-	63	-	63	-

Table II. The distribution of task temperatures across the groups of participants (Range indicates that more than one message was sought-after and these messages were of different temperatures)

3.3 Examining Recollections

We examined the participants' memories for each task at two levels. First, we investigated how able the participants were to remember if the information they needed was actually contained within their collections. Before each task was performed the participant was asked if he believed the information needed to solve the task was stored within his collection. The participant could answer yes, no, or not sure. This allowed us to determine at a high level the quality of the memory the users had for the information they were required to find and also their confidence in their recollection.

Second, we examined the participants' recollections in greater detail by employing a memory questionnaire. This is a technique that has been used often in the field of cognitive psychology to examine memory [see [Herrmann 1982] for a review]. After each task, the participant was questioned on what they were able to remember about the information they had just looked for, i.e. the information they had available to help them with the search. Firstly, the participants were asked about four attributes common to every email. They were asked if, before completing the task, they ***correctly*** remembered:

—Roughly when the email was sent
—The sender of the email
—What the email was about (we were interested in the topic of the email not the exact syntax of the subject line)
—The reason why the email was sent

Secondly, the participants were asked if they remembered four attributes that are only applicable to some emails. We refer to these as additional attributes. The participants were asked if, before completing the task, they ***correctly*** remembered:

—Any other person(s) who may have received the email (this could include both individual recipients as well as organised mailing lists)
—If the email had any attachments
—If the email contained an image
—If the email contained a link or url

To clarify, these questions related to what the participants remembered before completing the task, i.e. information that could have been used to guide their re-finding strategy. Naturally, there were differences in the quality of recollections for the various attributes. However, we used the rule of thumb that for the recollection to count it had to be potentially useful to the re-finding task. For example, remembering that an email was sent on "Tuesday 23rd January because it is my birthday" is different from remembering that "it was sent around Christmas time last year". Although, both of these recollections would be useful in a re-finding context. A recollection such as "I remember that it must have been some time in the last 3 years" is less useful and would probably not have been counted. However, the decision of whether or not a recollection should count was taken by the experimenter based on the context of the task and on the information provided by the participant. It should be noted, however, that in the vast majority of cases it was clear whether or not the participant remembered a useful attribute.

We asked about recollections retrospectively to ensure that no bias was exerted on the participants' behaviour while completing the tasks. There are advantages and disadvantages to employing this technique. We discuss these in section 6.

The information collected was analysed by establishing the percentages of tasks for which each attribute was remembered. Firstly, all of the tasks were analysed to determine an overall picture of the participants' recollection for email messages. Then, the data were analysed more closely by counting specific groups of tasks. This offered the opportunity to discover differences in the memories the participants had for different types of task, user and filing strategy.

4. MEMORY FOR EMAILS

The data collected are analysed and presented in two stages. First, in this section, we analyse the attributes that were remembered generally, as well as those remembered in the four scenarios outlined in section 2.3. We demonstrate how the remembered attributes changed in these situations and assert possible explanations for the changes. In a second step, in section 5 we try to validate our assertions based on the data. We present statistical models created from the data that illustrate how much influence different factors had on the attributes that were remembered.

4.1 Were participants able to remember if collections held the information they needed?

Before looking at the attributes that were remembered, we examined the data regarding whether or not the participants believed that the information needed to solve tasks was contained within their collections. Although this question was asked primarily to gauge the realism of the re-finding tasks – people would not search a location for information if they did not believe it was there to find – the answers that the participants gave and the way they provided the answers revealed something about their recollections for the information we asked them to find.

Overall, the data indicate that the participants were generally sure that the information that was required to solve the task was in their collections. In 159 (87.8%) of the 181 tasks, the participants were positive that the information required was contained within their collection. Only in 6 tasks (3.3%) were participants sure that the information was not contained within their collections and in 24 tasks (13.3%) they were unsure. As shown in Table III this pattern was reasonably consistent across the groups of participants. This suggests firstly, that the participants knew what information would help them complete the tasks and confirms that searching their email collections to find the information is a realistic scenario. Secondly, it suggests that participants have overall reasonably good recollections for the email messages that they receive and thirdly, that they seem to be confident in these recollections.

	P.Grad	%	U.Grad	%	Researcher	%	Total	%
Contains	57.0	90.5	54.0	85.7	48.0	76.2	159.0	84.1
Doesn't Contain	2.0	3.2	1.0	1.6	3.0	4.8	6.0	3.2
Not sure	4.0	6.4	8.0	12.7	12.0	19.1	24.0	12.7
Total	63	100	63	100	63	100	189	100

Table III. The numbers of tasks for which the required information was remembered to be in the participant's collection

The following subsections examine what attributes the participants tended to remember and demonstrate how the recollected attributes change in the scenarios discussed in section 2.3.

4.2 High-level recollections

Table IV shows the percentages of tasks for which participants remembered the various attributes. The percentages are given for the different types of task (lookup, item and multi-item) and temperatures (hot, warm, and cold).

The most frequently remembered attribute was the topic of the email, which was remembered in 85.1% of tasks. This was followed by the reason the email was sent (80.9%), the sender of the email (77.1%), and temporal information (57.5%). Other recipients were remembered in 46.8% of the tasks, links or URLs in 21.8% of the tasks, and attachments were remembered in 12.8% of the tasks. Images were reported as being remembered least often - only in 2.1% of the tasks did participants report remembering that the email contained an image. Nevertheless, it is likely that very few emails had attached images, and as there is no way of knowing what percentage of the tasks actually required the participants to retrieve

Task Type	When	Sender	Topic	Reason	Other recp.	Attach.	Image	Link	#Tasks
All tasks	57.5	77.1	85.1	80.9	46.8	12.8	2.1	21.8	188
Lookup	60.3	57.1	74.6	76.2	36.5	3.2	1.6	33.3	63
Item	58.7	68.3	85.7	87.3	42.9	6.4	0.0	6.4	63
Multi-item	34.9	90.5	82.5	66.7	46.0	14.3	4.8	11.1	63
Hot	71.1	95.6	91.1	88.9	53.3	22.2	2.2	17.8	45
Warm	68.9	71.1	93.3	91.1	44.4	17.8	2.2	37.8	45
Cold	56.7	56.7	80.0	81.7	41.7	6.7	3.3	26.7	60

Table IV. The percentages of all tasks in which the attributes were remembered

emails containing images, links or URLs, or attachments, it is difficult to determine the importance of these additional attributes. However, what can be said is that for some tasks, and in the case of other recipients, a fairly large percentage of tasks, the participants had access to these extra pieces of information to help them search. Further, because not all of the emails have the additional attributes, such recollections could be useful for re-finding because of their discriminative power.

Overall, the data indicate that for most tasks the participants reported remembering quite a lot about the email(s) they were looking for. In the vast majority of tasks the participants remembered what the email was about, why it was sent, as well as who sent it. Additionally, for many tasks these recollections were supplemented with additional temporal information, information about other recipients of email, as well as other attributes. Figure 2 shows this graphically, presenting the number of attributes common to all email messages that were remembered for different types of task as a boxplot. The boxplots are top heavy, indicating that for all of the task types the number of remembered attributes was high. Another way of considering this data is that in 42.3% of tasks, the participants remembered all of the common attributes, in 74.1% of the tasks 3 or more were remembered and in 85.9% of tasks participants remembered 2 or more common attributes. Therefore, as participants re-find based on what they can remember, the data suggest that for the majority of the assigned tasks, the participants had options regarding which attributes to use when re-finding.

4.3 Is there evidence of changing recollection as time goes by?

In this section we present analyses of the recollection data to determine if the length of time between accessing and re-accessing the required email(s) influenced how much or the kinds of attributes that the participants remembered. Table IV shows that time seems to have influenced what was remembered. In six of the eight attributes, the attribute was remembered less often for cold tasks than for warm tasks. The only attributes not to follow this trend were the image and link or URL attributes. In the case of the image attribute, there is probably not enough data to form a good pattern because images were only reported as being remembered in 4 out of the 189 tasks.

The trend for most attributes was a highest percentage remembered for hot tasks, followed by a lower percentage for warm tasks, followed by a lower still percentage for cold tasks. This is confirmed in Figure 3, which plots as a bar chart the percentages of tasks for which the various attributes were remembered. Figure 3 shows that the attributes were consistently remembered more often for hot tasks

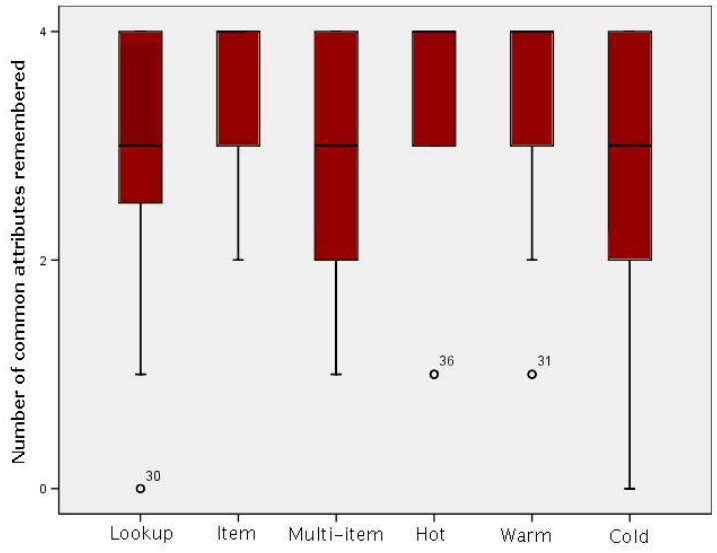

Fig. 2. Boxplot of the common attributes remembered for different task types

than for warm tasks and more often for warm tasks than for cold. Thus, the data suggest that recollections for email messages follows the same trends as recollection for other types of information, where over time there is a gradual degradation in memory.

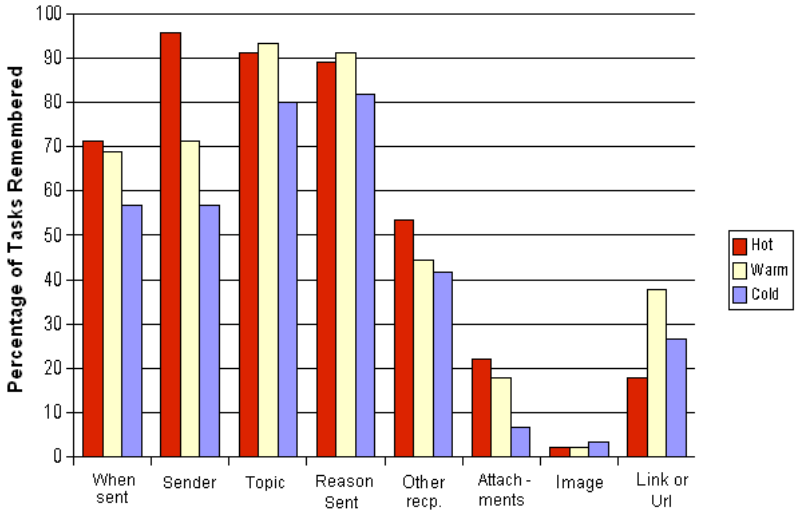

Fig. 3. Bar chart showing the percentage of tasks for which the various attributes were remembered and how this changed for tasks of different temperatures

Figure 3 also demonstrates that some attributes were more susceptible to forget-

ting over time than others. The temperature effect was most obvious on the sender attribute. In 95.6% of hot tasks the sender was remembered. This compares with 71.1% of warm tasks and only 56.7% of cold tasks. This means that as time goes by, participants are most likely to forget who sent an email. This finding suggests that re-finding tools should rely less on the sender attribute for older information.

The common attribute least affected by forgetting over time was the reason sent attribute. For hot tasks the reason sent was remembered in 88.9% of tasks, slightly more often for warm tasks (91.1%) and slightly less often for cold tasks (81.7%). These results suggest that a filing system organised by reason or topic would support the re-finding of older information more effectively than an organisation based on the sender of the messages.

4.4 Did the participants remember different things for different types of task?

This section analyses the recollection data across the three types of task completed: lookup tasks, item tasks and multi-item tasks. The data suggest that the participants remembered similar attributes for look-up and item tasks, but different attributes for multi-item tasks. This is shown graphically in Figure 4, which shows the number of tasks for which the different attributes were remembered as a bar chart and how the percentages varied depending on the type of task being performed. The most commonly remembered attributes for lookup and item tasks were the reason the email was sent, the topic of the email, when the email was sent, and the sender of the email.

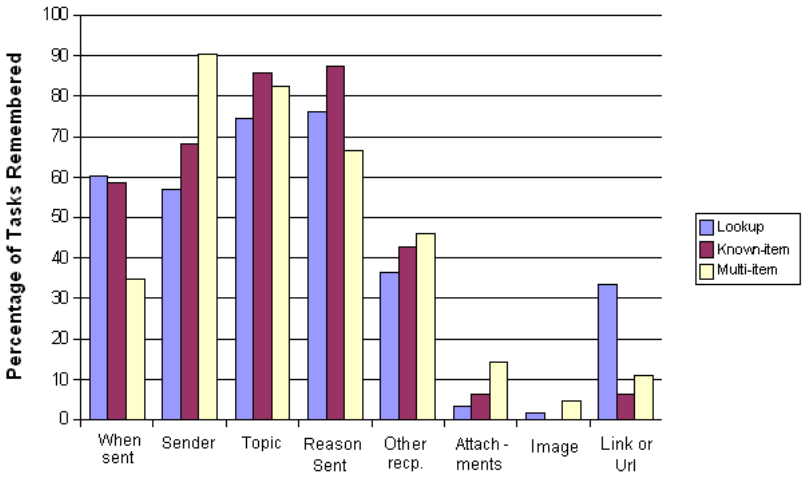

Fig. 4. Bar chart depicting the percentage of tasks for which the different attributes were remembered for different types of task

Multi-item searches had a different pattern of associated recollection. In multi-item tasks only three main attributes were commonly remembered: the sender of the email, the topic of the email and the reason the email was sent. One noticeable difference for multi-item tasks was that the participants had access to temporal

memories less often (34.9%) than in the other categories of tasks (60.3% for lookup tasks and 58.7% for item tasks). This makes sense because in multi-item tasks participants were often looking for emails or information within emails that were received at different time periods. In multi-item tasks, almost all of the time (90.5% of tasks) the participants remembered the sender of the email(s) that they were looking for. This is compared to far lower percentages of item tasks (68.3%) and lookup tasks (57.1%). Further, despite the similarities in the attributes remembered for lookup and item tasks, the participants tended to remember the attributes less often for lookup tasks than they did for item tasks. Again, Figure 4 confirms this graphically.

The changes between the recollections for the different types of tasks suggests that varying tool support may be required for different kinds of tasks.

4.5 Did different types of users remember different things?

In addition to examining the recollection data across different types of task, the data were analysed more closely to determine whether patterns existed in the recollections of the different classifications of user. As discussed above, our three categories of users – undergraduate students, postgraduate students, and researchers – used email for very different purposes, had different collection sizes and had been using their email accounts for different lengths of time. The data were analysed to determine if these properties affected what was remembered for the email tasks performed in this email re-finding study. Table V shows the percentages of tasks for which the various attributes were remembered by the three user groups.

It is important to acknowledge that when analysing the recollected attributes for the three groups that although the participants performed equal amounts of lookup, item, and multi-item tasks, the tasks performed by the different groups were not the same. Nevertheless, the tasks each group performed reflected the real tasks recorded by this group of participants in the task collection phase. Therefore, by comparing the attributes remembered by different groups of users, what we may actually be comparing is the recollections with respect to the different types of task that people perform and the collections that they have.

The data revealed differences between the attributes that the three groups of participants remembered. The postgraduate and undergraduate participants remembered similar types of attributes, on average in the following order of frequency: topic of the email, the sender, the reason the email was sent and then temporal information. The researchers tended to remember different attributes. They remembered other recipients most often, followed by the reason and the sender of the email, the topic and temporal information.

The data indicate that the researchers seemed to have generally remembered less than the other groups. This is shown graphically in Figure 5, which plots the percentages of tasks for which the various attributes were remembered as a bar chart. The fact that the researchers seemed to remember less about their emails is understandable considering that their collections were much larger than those of the participants in the other groups, suggesting that size of collection may effect what people remember. The only attribute that the researchers remembered more often than participants from the other groups was the "other recipient" attribute. This is very obvious in Figure 5. A plausible explanation for this is that many of

Group	Task Type	When	Sender	Topic	Reason	Other recp.	Attach.	Image	Link	#Tasks
Postgrad.	All tasks	61.9	81.0	95.2	77.8	33.3	12.7	0.0	22.2	63
	Lookup	71.4	47.6	95.2	90.5	33.3	4.8	0.0	38.1	21
	Item	81.0	100.0	100.0	100.0	47.6	33.3	0.0	28.6	21
	Multi-item	33.3	95.2	90.5	42.9	19.1	0.0	0.0	0.0	21
	Hot	100.0	100.0	100.0	84.6	53.9	23.1	0.0	15.4	13
	Warm	69.2	69.2	96.2	96.2	38.5	15.4	0.0	34.6	26
	Cold	50.0	50.0	87.5	75.0	0.0	0.0	0.0	37.5	8
Undergrad.	All tasks	57.1	77.8	88.9	92.1	30.2	20.6	4.8	36.5	63
	Lookup	47.6	66.7	71.4	76.2	9.5	0.0	0.0	47.6	21
	Item	66.7	76.2	95.2	100.0	38.1	23.8	0.0	28.6	21
	Multi-item	57.1	90.5	100.0	100.0	42.9	38.1	14.3	33.3	21
	Hot	56.5	100.0	95.7	100.0	43.5	26.1	0.0	21.7	23
	Warm	58.3	91.7	83.3	83.3	33.3	33.3	8.3	66.7	12
	Cold	60.0	56.0	84.0	88.0	20.0	12.0	8.0	40.0	25
Research	All tasks	53.2	72.6	71.0	72.6	77.4	4.8	1.6	6.5	62
	Lookup	61.9	57.1	57.1	61.9	66.7	4.8	4.8	14.3	21
	Item	85.0	75.0	100.0	100.0	90.0	5.0	0.0	5.0	20
	Multi-item	14.3	85.7	57.1	57.1	76.2	4.8	0.0	0.0	21
	Hot	66.7	77.8	66.7	66.7	77.8	11.1	11.1	11.1	9
	Warm	85.7	42.9	100.0	85.7	85.7	0.0	0.0	0.0	7
	Cold	55.6	63.0	74.1	77.8	74.1	3.7	0.0	11.1	27

Table V. The percentages of tasks for which the various attributes were remembered by different groups of users for different types of task

the emails they received revolved around groups of people. Previous work has noted that academics often think about research themes in terms of the people working in those areas and use people as an organisational thread to tie semantic themes together [Case 1991; Elsweiler et al. 2007]. Interviews with the participants in this email study seem to confirm this theory. The main uses of email for participants in the research group were research, teaching and social activities, all of which revolved around groups. It was often the case that the researchers had specific mailing lists that correspond to their activities e.g. members of their research group, department, social groups, groups of students taking a class etc. This is in contrast to the other groups who used their email account purely for emails that they felt applied to them.

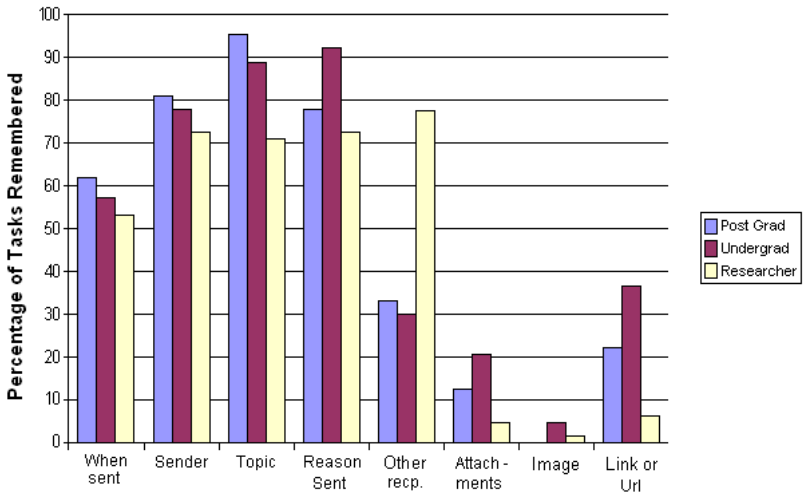

Fig. 5. Bar chart showing the percentages of tasks for which the various attributes were remembered and how this changed for the different groups of users

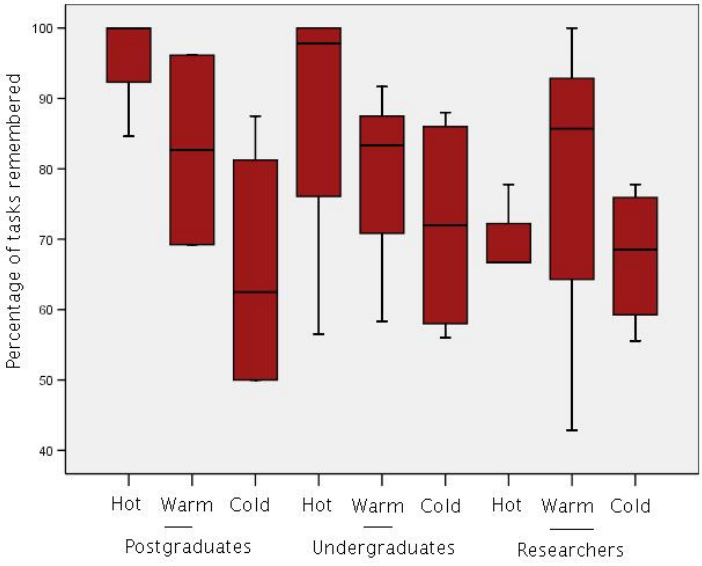

Fig. 6. A boxplot showing the percentages of tasks remembered by the different groups of participants for tasks of different temperatures

The researchers' poorer overall recollection is particularly noticeable for hot tasks. Whereas for hot tasks the undergraduate and postgraduate participants almost always had access to the sender of the email, what the email they were looking for was about and the reason it was sent, this was not the case for participants in the researchers group. For hot tasks, that is in a period of up to one week since last accessing the information, researchers were unable to remember the sender of the email in over 12% of tasks and unable to remember the topic of the email, the reason why the email was sent and any temporal information in one third of tasks.

The researchers did perform more cold tasks than the other groups. However, this is unlikely to be the explanation for their poorer overall recollection. Even though cold tasks were generally associated with poorer recollections than hot and warm tasks, this was not the case for the participants in the researchers group. Figure 6 shows that both the undergraduate and postgraduate participants remembered more for hot than warm tasks and more for warm tasks than cold. The researchers were the only group to break this temperature trend. The data demonstrate that the researchers remembered more for warm tasks than for both hot and cold. This finding suggests that even though the researchers performed more cold tasks than the other groups, this is unlikely to be the reason that the researchers remembered less generally.

To summarise, the data indicate that the recollections differed across the three groups of participants. The researchers seemed to remember different properties about their emails than the post and undergraduate student groups – usually remembering less. However, the researchers were able to remember other recipients of an email more often than the other groups.

4.6 Did the filing strategy employed influence what the participants remembered?

This section examines the recollection data with respect to the three filing groups: no-filers, frequent filers, and spring-cleaners. Only the undergraduate participants and the researchers were included in these analyses. The postgraduate group was discounted because almost the entire group employed a no-filing strategy and including them would have imbalanced the study. There was a reasonable balance in the filing strategies employed by the undergraduate and researcher participants [see Table I], meaning that a fair comparison could be made. Table VI shows the email properties of the groups of users employing different filing strategies.

Property	No-Filers	Freq.Filers	spring-cleaners
Number of emails (mean, std dev)	2083.67, 2607.48	2723.67,2229.44	3818.00, 4560*
Number in group	4	5	5
Emails received per day (median, IQR)	1.06, 12.7	2.03, 9.82	2.66, 7.11*
Experience (median, IQR)	4	3.5	4
Avg. Number of hot tasks per participant	2.17	2	2.6
Avg. Number of warm tasks per participant	1.17	2	1.2
Avg. Number of cold tasks per participant	3.83	3	3.8

Table VI. The email properties of the different filing groups (statistically significant values marked by *)

Strategy	Task Type	When	Sender	Topic	Reason	OR	Attach.	Image	Link	#Tasks	%tasks
No-Filers	All tasks	63.0	81.5	85.2	88.9	53.7	20.4	5.6	27.8	54	100.0
	Lookup	55.6	72.2	66.7	77.8	27.8	0.0	0.0	33.3	18	33.3
	Item	88.9	88.9	100.0	100.0	61.1	22.2	0.0	16.7	18	33.3
	Multi-item	44.4	83.3	88.9	88.9	72.2	38.9	16.7	33.3	18	33.3
	Hot	46.2	100.0	92.3	92.3	53.9	23.1	0.0	38.5	13	24.1
	Warm	85.7	85.7	85.7	85.7	57.1	42.9	14.3	57.1	7	13.0
	Cold	73.9	73.9	82.6	91.3	47.8	17.4	8.7	26.1	23	42.6
Frequent-Filers	All tasks	34.6	69.2	76.9	69.2	50.0	3.9	0.0	11.5	26	100.0
	Lookup	22.2	66.7	44.4	33.3	33.3	0.0	0.0	22.2	9	34.6
	Item	50	62.5	100	100	62.5	12.5	0	12.5	8	30.7
	Multi-item	33.3	77.8	88.9	77.8	55.6	0.0	0.0	0.0	9	34.6
	Hot	66.7	83.3	83.3	83.3	33.3	16.7	0.0	0.0	6	23.1
	Warm	33.3	83.3	83.3	66.7	33.3	0.0	0.0	33.3	6	23.1
	Cold	22.2	44.4	66.7	55.6	55.6	0.0	0.0	11.1	9	34.6
spring-cleaners	All tasks	57.8	71.1	75.6	82.2	55.6	8.9	2.2	20.0	45	100.0
	Lookup	73.3	46.7	73.3	80.0	53.3	6.7	6.7	33.3	15	33.3
	Item	73.3	66.7	93.3	100.0	66.7	6.7	0.0	20.0	15	33.3
	Multi-item	26.7	100.0	60.0	66.7	46.7	13.3	0.0	6.7	15	33.3
	Hot	69.2	92.3	84.6	92.3	61.5	23.1	7.7	7.7	13	28.9
	Warm	83.3	50.0	100.0	100.0	66.7	16.7	0.0	33.3	6	13.3
	Cold	57.9	52.6	84.2	89.5	47.4	0.0	0.0	31.6	19	42.2

Table VII. The percentages of tasks for which the various attributes were remembered and how this varied for participants utilising different filing strategies

The data show that, of the three filing groups, the group that remembered most were the no-filers. On average, the no-filers remembered more than the spring-cleaners and the spring-cleaners remembered more than the filers. When the recollections are analysed for the different attributes [see Figure 7], nearly every attribute follows this pattern. The only attribute that did not was the "other recipients" attribute. In this case all three groups seemed to remember with similar frequency (no-filers = 53.7%, spring-cleaners = 55.6% and frequent filers =50.0%), but again the frequent filers had the lowest percentage.

Another pattern observed in the data was that the participants who employed a frequent filing strategy had access to temporal information very infrequently

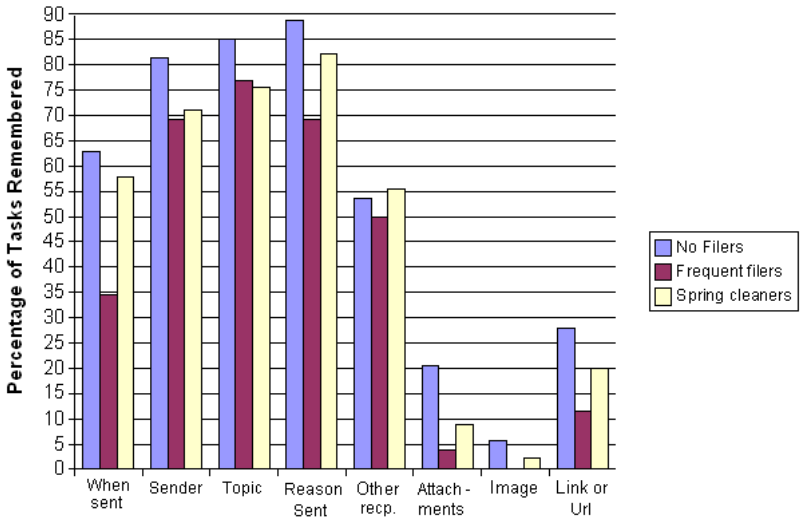

Fig. 7. Bar chart showing the percentages of tasks for which the various attributes were remembered and how this changed for the different filing groups

compared to the other groups. This is clearly shown in Figure 7. Thus, the data indicate that the filing strategy that the participants employed influenced what and how much they remembered about their emails. The evidence seems to favour Mander's argument that employing a no filing strategy leads to remembering more about the information in one's collection.

This section has examined what the participants remembered in the scenarios outlined by our review of psychology literature in section 2.3. It has been demonstrated that the attributes that our participants remembered seemed to change depending on the temperature of the task, the type of task, the type of user and the filing strategy of the user. However, the problem with the approach we have taken so far is that there are other variables that may be influencing the data. In the following section we examine the data statistically to account for these variables and determine the factors that had the greatest influence on whether or not the participants remembered particular attributes.

5. FACTORS INFLUENCING WHETHER THE ATTRIBUTES WERE REMEMBERED

Four logistic regression models were developed to determine whether any factors were associated with each of the four principal memory attributes; when sent, sender, topic and reason sent. All available factors (19 in total) collated from the user study were analysed initially using a stepwise procedure in order to isolate any significant relationships. The stepwise procedure enters and removes factors at each step assessing the overall goodness of fit of the linear regression model [McCullagh and Nelder [1989] provide an overview on generalized linear models and the stepwise procedure].

Following this initial investigation, we considered specific factors of interest (collection size, user group, temperature, experience, task type and filing strategy)

alongside those factors included in the model during the stepwise procedure. The reasoning behind this decision was that automatic model building procedures can be erroneous as they do not consider the real world importance of each factor. Therefore, the final models presented for each memory attribute also included those factors believed to be important prior to investigation (based on our analyses in section 4). These factors were entered into the model to assess both their effect (if any) on the memory attribute in question and also the relationship between these factors and the other remaining variables found to be significant. The models are presented in Tables VIII to XI and the effect plots for the models are shown in Figures 8 to 10[4]. For each model, the regression coefficient, p-value, odds ratio and associated 95% confidence interval are presented for all factors included. In the effect plots the solid line represents the probability of a user from the sample answering yes irrespective of the factor. Therefore, if the factor level is above the line this represents a positive effect and vice versa.

Factor	Level	Coef	SE Coef	p-value	OR	95% CI
Constant		-3.6	1.3	<0.01		
Experience		**0.62**	**0.26**	**0.02**	**1.86**	**[1.12, 3.10]***
Temperature	Warm	0.49	0.47	0.29	1.63	[0.65, 4.11]
	Hot	0.79	0.45	0.08	2.19	[0.90, 5.35]
	Range	**-1.54**	**0.5**	**<0.01**	**0.22**	**[0.08, 0.57]***
Filing Group	**No-filers**	**1.42**	**0.56**	**0.01**	**4.13**	**[1.39, 12.26]***
	spring-cleaners	**1.41**	**0.55**	**0.01**	**4.08**	**[1.38, 12.02]***
	No-filers vs spring-cleaners	-0.01	0.38	0.97	0.99	[0.47, 2.09]
Num emails		0	0	0.28	1	[1.00, 1.00]
Participant Group	PostGrad	1	0.55	0.07	2.72	[0.93, 7.98]
	Researcher	0.78	0.48	0.1	2.19	[0.86, 5.58]

Table VIII. Regression Model for "when sent" attribute (significant factors shown in bold and marked with *)

Factor	Level	Coef	SE Coef	p-value	OR	95% CI
Constant		0.66	1.48	0.66		
Num emails		**<0.01**	**<0.01**	**<0.05**	**1.00**	**[1.00,1.00]***
Experience		-0.01	0.30	0.98	0.99	[0.56, 1.77]
Temperature	Warm	0.35	0.48	0.47	1.42	[0.55, 3.65]
	Hot	**2.69**	**0.78**	**<0.01**	**14.68**	**[3.16, 68.32]***
	Range	**1.85**	**0.62**	**<0.01**	**6.36**	**[1.88, 21.54]***
Filing Group	No-filers	0.19	0.65	0.77	1.21	[0.34, 4.28]
	spring-cleaners	0.06	0.62	0.93	1.06	[0.31, 3.57]
Participant Group	PostGrad	-0.19	0.62	0.76	0.83	[0.25,2.77]
	Researcher	0.28	0.56	0.62	1.32	[0.44,3.98]

Table IX. Regression Model for "sender" attribute (significant factors shown in bold and marked with *)

The analyses in section 4.3 indicated that the participants tended to demonstrate poorer recollection when the information they were looking for had not been accessed for longer time periods. The models in this section show that even when other factors are considered, the temperature of the task still had an influence on whether or not attributes were remembered, with this variable featuring in the

[4]we have only displayed significant factors graphically

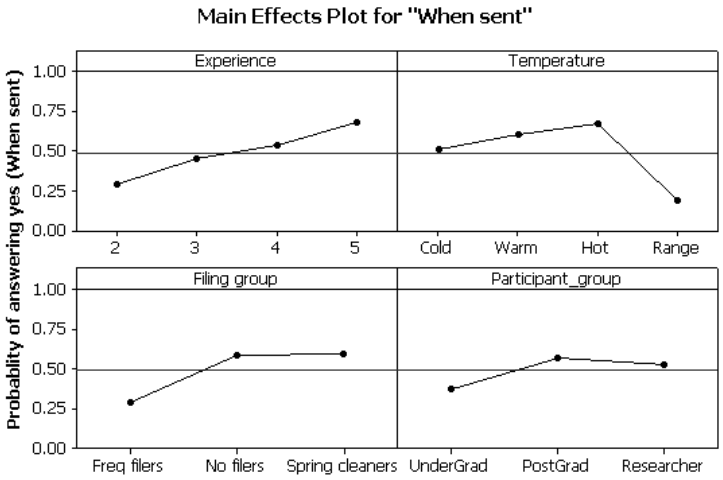

Fig. 8. Main effects plot for "When sent" attribute

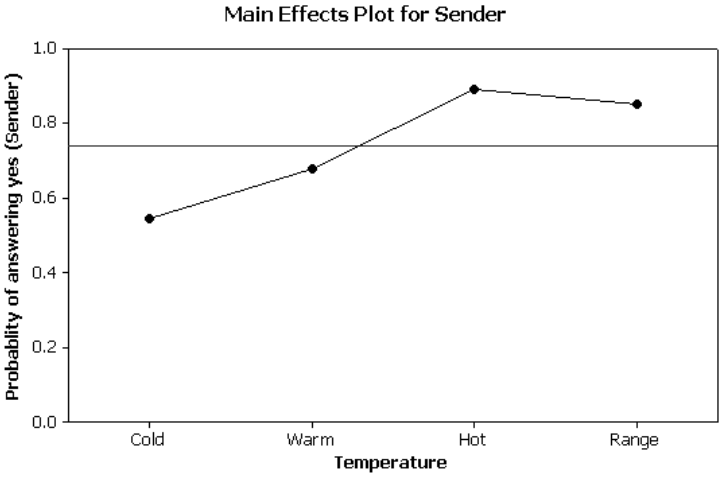

Fig. 9. Main effects plot for "sender" attribute

models for the "when sent", "sender" and "reason sent" attributes. However, it was only in the "sender" model that the influence of the task temperature was significant ($p<0.01$, OR=14.68, 95%CI= [3.16, 68.32] from Table IX)[5]. The other models had significant values for the variable, but only on the range category which was associated with multi-item tasks. This makes sense because when completing

[5]To help the reader understand the models we have supplied the p-value, the odds ratio and the confidence interval for this first example. However, to make the text easier to read we shall omit this information in the following cases. Referring instead only to the appropriate table

Factor	Level	Coef	SE Coef	p-value	OR	95% CI
Constant		2.13	2.18	0.33		
Temperature	Warm	0.27	0.78	0.73	1.31	[0.28, 6.01]
	Hot	0.89	0.75	0.24	2.42	[0.55, 10.58]
	Range	-0.87	0.65	0.19	0.42	[0.12, 1.51]
Filing Group	No-filers	0.93	0.82	0.25	2.53	[0.51, 12.52]
	spring-cleaners	-0.39	0.67	0.56	0.67	[0.18, 2.51]
Participant Group	PostGrad	0.58	0.92	0.53	1.79	[0.29, 10.91]
	Researcher	-0.26	0.80	0.75	0.77	[0.16, 3.68]
Experience		-0.31	0.45	0.49	0.73	[0.30.1.77]
Time in Secs		**0.01**	**<0.01**	**0.03**	**1.01**	**[1.00, 1.02]***
Num emails		>-0.01	<0.01	0.49	1.00	[1.00, 1.00]

Table X. Regression Model for "Topic" Attribute (significant factors shown in bold and marked with *)

Factor	Level	Coef	SE Coef	p-value	OR	95% CI
Constant		-0.56	1.84	0.76		
Preferred re-finding strategy	Browsing	0.88	0.67	0.19	2.41	[0.65, 8.88]
	No pref	**1.89**	**0.73**	**0.01**	**6.61**	**[1.57,27.87]***
Num emails		>-0.01	<0.01	0.99	1.00	[1.00, 1.00]
Experience		0.23	0.33	0.49	1.25	[0.66, 2.39]
Filing Group	**No-filers**	**1.63**	**0.74**	**0.03**	**5.09**	**[1.20, 21.59]***
	spring-cleaners	**1.76**	**0.73**	**0.02**	**5.79**	**[1.39, 24.19]***
Temperature	Warm	0.96	0.70	0.17	2.61	[0.67,10.21]
	Hot	0.46	0.63	0.47	1.58	[0.46, 5.39]
	Range	**-1.23**	**0.53**	**0.02**	**0.29**	**[0.10, 0.83]***
Participant Group	PostGrad	-0.99	0.70	0.17	2.61	[0.67, 10.21]
	Researcher	-0.76	0.70	0.28	0.47	[0.12, 1.84]

Table XI. Regression Model for "Reason sent" attribute (significant factors shown in bold and marked with *)

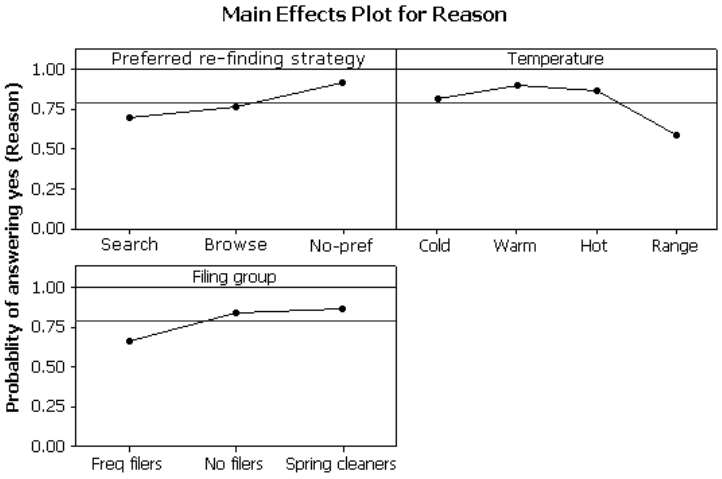

Fig. 10. Main effects plot for "reason sent" attribute

tasks that were categorised as "range", there was no precise date to remember as the messages containing the important information were received at different periods of time. Further, as the "reason sent"model shows, multi-item tasks required messages to be found that were sent for different reasons. Nevertheless, the models show that the more time elapsed between accessing and re-accessing an email, the less likely the participants were to remember the sender of the email.

The analyses in section 4.5 demonstrated differences in the patterns of recollection across the three participant groups in the study. However, the user group variable only featured in one of the logistic regression models ("when sent") and even in this model its influence was not significant (p=0.07). In section 4.5 we suggested that the size of the researchers' collections could be one explanation for the different recollections offered by the participant groups. There is evidence for this in the models, but only for the "sender" attribute (p=0.045). This indicates that participants with larger collections were statistically less likely to remember the sender of an email. Further, also relating to the properties of the different filing groups, the model for the "when sent" attribute shows that the experience of the user was a significant factor in determining whether this attribute would be remembered (p=0.02). According to the data, the participants with greater experience were more likely to remember when a sought-after email was sent. Thus, the data show that some of the properties that the different user groups exhibited affected what they were able to remember. Participants with larger collections were less likely to remember who sent an email, while participants with greater re-finding experience were more likely to remember when an email was sent.

The models also confirm the influence that the participants' filing strategy had on what they were able to remember. Both the "when sent" (p=0.01) and "reason sent"(p=0.02) models show that the filing strategy employed significantly influenced the probability that these attributes would be remembered. According to these models, participants who reported filing their emails regularly were less likely to remember when an email was sent or the reason that the email was sent. However, the models show no difference in the recollections of the participants with no-filing and spring-cleaning strategies.

The only factor from our analyses in section 4 that did not feature in the models was the type of task performed. This factor was out-weighed by the others in the models. However, we did find another factor that was not identified in our previous analyses. The "reason sent" model shows that the preferred form of re-finding had an effect on what was recalled (p=0.01). This model shows that participants who reported preferring to use a search interface when re-finding were less likely to remember the reason that an email was sent than those who reported no preference for either searching or browsing. This suggests that the finding strategy that a user employs impacts on what he will remember about his emails.

To summarise, this section has presented logistic regression models for each of the four principal remembered attributes. The models show that task temperature, re-finding experience, size of collection, filing strategy and preferred re-finding technique all influenced the attributes that the participants remembered.

6. LIMITATIONS

Before discussing the implications of our findings, it is important to acknowledge the limitations of the work. Our study relates to memory for email re-finding and the findings should only be considered in this context. A limitation of the work is that our study population consisted purely of computer scientists and computer science students and investigated their memories for tasks that revolved around work-based activities. Although computer scientists may not be representative of all email users, we argue that our results are generalisable, at least to some extent, for a number of reasons. Firstly, we included a group of users (the postgraduates) who had only recently started their course and who did not have a computer science background. These participants mainly had low levels of computer literacy and limited experience with email. Secondly, although we only examined re-finding tasks that were based on work and not leisure activities, we did examine each of the three categories of email re-finding tasks that previous work had shown users to complete in both work and leisure scenarios [Elsweiler and Ruthven 2007]. Also relating to the demographics of the study population, as mentioned in section 3.1, the participants in our study had far fewer emails in their collections than had been reported in previous studies. Nevertheless, we do not feel this detracts from the usefulness of our findings because regardless of the quantities of emails in the collections, the study analysed **real** users, performing **realistic** re-finding tasks, on their **own** collections.

Regarding the creation of experimental tasks, great care was taken to learn about the contents of the participants' collections as well as the kind of re-finding tasks they perform. The process involved recording real tasks that users in these groups performed and using these tasks as a template to create experimental tasks. It should be noted, however, that when asking a participant to perform a re-finding task it is necessary to tell the participant something about the information he should find before he can re-find it. This will undoubtedly affect the findings. However, again, great care was taken in the wording of the experimental tasks to minimise the effect. For example, rather than using phraseology that formed part of the textual content of the email, we chose wordings that would accurately communicate the information need, without providing keywords in the email text. For example, in task A2, rather than ask the participants to find information about the "MSDN academic alliance", we asked them to find information about how they would go about getting free software from Microsoft through the university. Of course we were not always privy to the wording of emails, so in many cases we created a context where information would be required without mentioning details about the email e.g. task A3. We also tried to limit the amount of named entities in the task descriptions. Further, the tasks were read aloud to the participants so that they were not assisted with spelling, nor could they refer back to the text again during the task.

Another limitation of our work is the number of variables present in the study. As noted above, other researchers have acknowledged the difficulties in performing PIM evaluations and one of the main challenges is controlling the variables present in experimental designs. In this study we have attempted to control the variables as much as possible through rotating the types of task performed and the experimental

system used around the participants. However, we concede that there are many variables that we were not able to control including the temperature of the task, the frequency with which the users perform that kind of task, the filing strategy of the users, the collection size of the user, and the difficulty of the task. Nevertheless, the way that the data were analysed, particularly in section 5, accounted for the uncontrolled variables, allowing us to isolate the factors that had an influence on the attributes that the participants remembered.

We would also like to mention limitations regarding the methods which were used to establish what the participants remembered. Firstly, we discuss the limitations with regards to our high-level analysis where we asked participants if they could remember if the information they required was in their collection. As it turned out, the majority of the assigned tasks involved finding information that **was** in the participants' collections. This may have added a bias to the findings because, as the required information was generally in the collections this may have led to the participants guessing that all of the tasks were such. However, we do not think the results were particularly biased and tell us something useful about what the participants remembered. We make this claim for a number of reasons. Firstly, the participants were not informed about how the tasks were created. All they knew was that the experimenter had no prior access to their email collections to create tasks and could not be sure that any information was in their collection. Secondly, many of the tasks made reference to an information need and did not specify what email would solve this need, relying instead on the participant making the connection between the task and the email(s) required. Indeed, sometimes different participants solved the same task using different email(s). Thirdly, we as experimenters had no idea if the information would actually be in the collections the participant may have chosen to delete the particular message that solved the task. It was surprising to us that this situation only occurred once, when one of the participant in the researchers group had deleted the email needed to solve the task. A last point we would like mention is that the main outcome of this analysis, that users had good recollection for their emails, was endorsed by later analyses, as well as the participants' behaviour while completing tasks.

Regarding the methodology used to discover the attributes of email messages that were remembered, we asked the participants what they were able to remember retrospectively, after they completed each task. The reason for this was that our experiment was primarily designed to be a system evaluation and we did not wish to influence the participants' re-finding behaviour by asking them about their memories before they performed each task. However, this means that we did not record what the participants remembered, rather what they thought they remembered. It also means that the process of re-finding and the information that they saw during the completion of the task may have subconsciously influenced what the participants believed they remembered. Again, however, we took steps to address this potential bias. During the completion of the task the participant voiced their thoughts aloud. This gave the experimenter some idea of what was remembered before the task and what was learned (or cued) during the task. In the retrospective questioning the experimenter was able to check with the participant when doubt arose. Nevertheless, the participants were generally good at determining what in-

formation they remembered and what information was cued and this information was communicated as part of the flow of conversation between the experimenter and the participant.

7. DISCUSSION

In this article we have presented a study of the recollections people have when re-finding email messages. We studied memories by using a memory questionnaire – a recognised technique from the field of cognitive psychology – while participants performed assigned tasks on their own collections. The aim of the study was to understand the role that memory plays in email re-finding and to better inform the design of email re-finding tools.

There were several outcomes to our work. It was discovered that:

—the participants generally demonstrated good abilities to remember their emails.
—the most frequently remembered attributes were semantic-based (the topic of the email and the reason the email was sent)
—the attributes that tended to be remembered changed in different scenarios
—the temperature of the task, the filing strategy of the participant, the participant's experience, the size of their collection, and their preferred method of re-finding all had an influence on the attributes that were remembered.

Below we summarise these findings and discuss what they mean in the context of designing re-finding systems and in terms of future work.

Overall, the data indicate that the participants generally had quite good recollections of their email messages and the information they were required to find. In section 4.1 it was discovered that the participants generally remembered whether or not the information they needed to solve the task was contained within their collection. This is an important finding because it means that the participants were looking for information that they would conceivably search their emails for and that they would have been able to recognise the information when they found it. The fact that they remembered which information would solve the information needs we supplied also hints at the capabilities of recollection for emails. It was also noted in section 4.2 that the participants typically remembered multiple attributes about the messages they were required to find. This means, in theory at least, that in the majority of tasks, they would have had options regarding how to approach the re-find task.

Despite these positive indicators about the participants' overall recollection, there were also situations where the participants remembered less. Therefore, it is important that re-finding tools account for this and support the attributes that are likely to be remembered. This finding is similar to that of Kalnikaité and Whittaker [2007] who examined how people use memory prosthetics and discovered that for memory prosthetics to be successful they need to work in synergy with the user's memory. According to our results the most frequently remembered attributes were semantic-based. The topic of the email was remembered in 85.1% of tasks messages and the reason the email was sent was remembered in 80.9% of tasks. This suggests that organisations based on the topic of would best support the recollections of our participants. These findings also seem to endorse task-based [Gwizdka 2002;

Bellotti et al. 2003] and project-based [Jones et al. 2005] approaches, rather than temporal [Ringel et al. 2003; Freeman and Gelernter 1996; Lansdale and Edmonds 1992] or people-based approaches [Whittaker et al. 2002]. Our data indicate that focusing support on temporal memories is probably unwise for email re-finding. The participants were only able to remember when an email was sent in 57.45% of the tasks and our analyses in section 5 show that the ability to remember when an email was sent was related to experience. This means that temporal-based approaches are particularly inappropriate for inexperienced users. Similarly our findings discourage organisations based primarily on the sender of the email. Not only was the sender attribute remembered less frequently than the semantic attributes, but our analyses show that recollections of the sender were correlated to the age of the information and the size of collection. Thus, as collection sizes increase and the more time had elapsed since an email was last read, the less likely the user will be to remember the sender of the message. No such relationships were discovered for the other attributes.

The finding that the participants tended to remember different attributes in different situations, however, suggests that relying on one form of memory or one particular organisation may not be the best approach. In section 4 the analyses showed that the participants exhibited different patterns of recollection when completing different types of task, when looking for information of different temperatures, and when they used different filing strategies. This finding advocates flexibility in re-finding tools, where the tool can support the different patterns of recollection associated with various situations. It also endorses the approaches of groups, such as Dumais et al. [2003], Cutrell et al. [2006], and Elsweiler et al. [2007] who have tried different approaches to increasing the flexibility of re-finding tools.

Another, slightly surprising finding was the impact of filing strategy on what the participants remembered. The analyses in section 4.6 suggested and those in section 5 confirmed that utilising a frequent filing strategy made the participants less likely to remember attributes about their email messages. This seems to endorse Mander's argument about the reminding function that browsing one's collection performs. The frequent filers seemed to exhibit particularly poor recollections for temporal information. Mander's theory can be extended to offer a potential explanation for this. As emails in the inbox are presented as a temporally ordered list, when spring-cleaners and no-filers interact with their messages residing in their inboxes, they are reminded of the time they were received because they see them in the context of other surrounding messages. Filers, on the other hand, remove this context when they move their messages to other folders. This means that not only do they come in to contact with filed messages less often, but when they do interact with the messages, there is no reminder of the temporal context. Nevertheless, this seems to be a rather paradoxical situation, where those who exert the most effort to keep their information organised actually exhibit the poorest recollection for their emails. However, just because a person remembers more by applying a particular filing strategy, it does not necessarily follow that this is the most effective strategy to use. It is possible that the participants who employed a frequent filing strategy are required to remember less in order to re-find because of the quality of their organisations. Previous work seems to indicate that this might be the case. For

example, the work of Case [1991], Cole [1982] and Kwasnik [1989] all show that people make classificatory choices based on the topic and purpose of documents – the kinds of attributes that our study show that people remember. However, it would be interesting to examine how this impacts on their re-finding performance. We will examine the relationship between recollection and re-finding performance in our data in a future article. Nevertheless, our findings here suggest that a simple addition to traditional folder-based interfaces may allow filers to experience the benefits of filing, in addition to the increased recollection associated with a no-filing approach. If the user had an option to toggle to a view that showed filed emails in their original context in the inbox, perhaps colour-coded to indicate that they had been filed in a particular folder, this may assist their recollection by reminding the user about the email as well as providing a temporal context. We are currently investigating the use of interfaces with this feature and will report on our findings in a future article.

A final point that we wish to discuss relates to the cueing of recollections. In our study it was obvious that the participants remembered far more details about the sought-after information than they actually used in the process of re-finding. Similarly, both Blanc-Brude and Scapin [2007] and Gonçalves and Jorge [2004] reported enhanced recollection abilities when cues were provided. This indicates that there may be benefit in employing systems that prompt the participant to use particular attributes when re-finding. Cues have been explored in the past, for example in the work of Lansdale and Simpson [1990] and in Elsweiler et al. [2007]. However, it would also be interesting to determine how systems that prompt for different attributes alter the re-finding performance of participants. For example, would systems that prompt participants to use more discriminative attributes offer quicker or more efficient re-finding? This could be one future avenue of exploration.

8. CONCLUSIONS AND FUTURE WORK

This article has investigated the attributes that people tend to remember when re-finding email messages. We demonstrated that although people generally remember quite a lot about their emails, there are situations in which people remember less and in these situations it may be more difficult to re-find the information required with existing tools. We showed that several factors influence how much and what attributes are remembered including: the time that has elapsed since the information was accessed, the experience of the user, the number of emails in the collection and the user's filing strategy. We also outlined some of the implications the findings have for the design of re-finding tools.

To build on this work, what is needed is an understanding about how recollection relates to re-finding performance and how different types of re-finding tools support the patterns of recollection observed here. Answers to these questions would provide good insight into how new re-finding tools should be designed. In a follow-up article, we analyse the findings of this study with these aims in mind. First we examine the performance data irrespective of the systems used to establish the importance of memory in re-finding. Then, we explore how each of the three systems used facilitated re-finding of information based on what the participants remembered and look at which features of the interfaces assisted in which situations and why.

REFERENCES

Baddeley, A. 1966. Short-term memory for word sequences as a function of acoustic, semantic and visual similarity. *Quarterly Journal of Experimental Psychology 18*, 362–365.

Bälter, O. 2000. Keystroke level analysis of email message organization. In *CHI '00: Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM Press, New York, NY, USA, 105–112.

Barreau, D. K. and Nardi, B. 1995. Finding and reminding: File organization from the desktop. *ACM SIGCHI Bulletin 27,* 3, 39–43.

Bellotti, V., Ducheneaut, N., Howard, M., and Smith, I. 2003. Taking email to task: the design and evaluation of a task management centered email tool. In *CHI '03: Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM, New York, NY, USA, 345–352.

Blanc-Brude, T. and Scapin, D. L. 2007. What do people recall about their documents?: implications for desktop search tools. In *IUI '07: Proceedings of the 12th international conference on Intelligent user interfaces*. ACM, New York, NY, USA, 102–111.

Boardman, R. 2004. Improving tool support for personal information management. Ph.D. thesis, Imperial College London.

Boardman, R., Spence, R., and Sasse, M. A. 2003. Too many hierarchies? the daily struggle for control of the workspace. In *Proceedings of HCI International 2003*, C. Jacko, J.Stephanidis, Ed. Vol. 1. 616–620.

Bower, G., Clark, M., Lesgold, A., and Winzenz, D. 1969. Hierarchical retrieval schemes in recall of categorized word lists. *Journal of Verbal Learning and Verbal Behavior 8*, 323–343.

Brown, N. R., Rips, L. J., and Shevell, S. K. 1985. The subjective dates of natural events in very long term memory. *Cognitive Psychology, 17*, 139–177.

Capra, R. G. and Perez-Quinones, M. A. 2005. Using web search engines to find and refind information. *Computer 38,* 10, 36–42.

Capra, R. G. and Perez-Quinones, M. A. 2006. Factors and evaluation of refinding behaviors. In *SIGIR 2006 Workshop on Personal Information Management, August 10-11, 2006, Seattle, Washington*.

Carroll, J. 1982. Creative names for personal files in an interactive computing environment. *International Journal of Man-Machine Studies 16*, 405–438.

Case, D. O. 1991. Conceptual organization and retrieval of text by historians: The role of memory and metaphor. *JASIST 42,* 9, 657–668.

Clark, H. H. and Clark, E. V. 1977. *Psychology and language*. New York: Harcourt Brace & Jovanovich.

Cohen, G. 2004. *Memory in the Real World (2nd Edition)*, (2nd Edition) ed. Psychology Press.

Cole, I. 1982. Human aspects of office filing: Implications for the electronic office. In *Proceedings Human Factors Society*.

Craik, F. and Lockhart, R. 1972. Levels of processing: A framework for memory research. *Journal of Verbal Learning & Verbal Behavior 11*, 671–684.

Cutrell, E., D.Robbins, S.Dumais, and R.Sarin. 2006. Fast, flexible filtering with phlat. In *CHI '06: Proceedings of the SIGCHI conference on Human Factors in computing systems*. ACM Press, New York, NY, USA, 261–270.

Czerwinski, M. and Horvitz, E. 2002. An Investigation of Memory for Daily Computing Events. In *Proceedings of HCI 2002*, S. Brewster and M. Zajicek, Eds. 230–245.

D'Andrade, R. G. 1995. *The Development of Cognitive Anthropology*. New York: Cambridge University Press.

Dourish, P., Edwards, W. K., LaMarca, A., Lamping, J., Petersen, K., Salisbury, M., Terry, D. B., and Thornton, J. 2000. Extending document management systems with user-specific active properties. *ACM Trans. Inf. Syst. 18,* 2, 140–170.

Ducheneaut, N. and Bellotti, V. 2001. E-mail as habitat: an exploration of embedded personal information management. *interactions 8,* 5, 30–38.

Dumais, S., Cutrell, E., Cadiz, J., Jancke, G., Sarin, R., and Robbins, D. 2003. Stuff i've seen: a system for personal information retrieval and re-use. In *SIGIR '03: Proc. 26th annual international ACM SIGIR conference on Research and development in information retrieval*, M. Sanderson, Ed. ACM Press, New York, NY, USA, 72–79.

Ebbinghaus, H. 1885. *ber das Gedchtnis. Untersuchungen zur experimentellen Psychologie (Memory: A Contribution to Experimental Psychology)*. Leipzig: Duncker and Humblot.

Elsweiler, D. and Ruthven, I. 2007. Towards task based pim evaluations. In *Proceedings of SIGIR*.

Elsweiler, D., Ruthven, I., and Jones, C. 2007. Towards memory supporting personal information management tools. *J. Am. Soc. Inf. Sci. Technol. 58,* 7, 924–946.

Eysenck, M. 1992. *Anxiety: The Cognitive Perspective*. Lawrence Erlbaum.

Fisher, D., Brush, A. J., Gleave, E., and Smith, M. A. 2006. Revisiting whittaker & sidner's "email overload" ten years later. In *CSCW '06: Proceedings of the 2006 20th anniversary conference on Computer supported cooperative work*. ACM, New York, NY, USA, 309–312.

Freeman, E. and Gelernter, D. 1996. Lifestreams: a storage model for personal data. *SIGMOD Record (ACM Special Interest Group on Management of Data) 25,* 1, 80–86.

Gonçalves, D. and Jorge, J. A. 2004. Describing documents: what can users tell us? In *IUI '04: Proceedings of the 9th international conference on Intelligent user interfaces*. ACM, New York, NY, USA, 247–249.

Grossman, L. and Eagle, M. 1970. Synonymity, antonymity, and association in false recognition responses. *Journal of Experimental Psychology 83*, 244–248.

Gwizdka, J. 2002. Reinventing the inbox: supporting the management of pending tasks in email. In *CHI '02: CHI '02 extended abstracts on Human factors in computing systems*. ACM, New York, NY, USA, 550–551.

Hayes, P. J., Andersen, P. M., Nirenburg, I. B., and Schmandt, L. M. 1990. Tcs: a shell for content-based text categorization. In *Proceedings of the sixth conference on Artificial intelligence applications*, M. Fox, Ed. IEEE Press, Piscataway, NJ, USA, 320–326.

Heesch, D. and Rüger, S. 2004. Nnk networks for content-based image retrieval. In *26th European conference on infromation retrieval (ECIR 2004), Sunderland, England*, S. McDonald and J.Tait, Eds. Number 2997 in LNCS. Springer-Verlag,, 253 266.

Herrmann, D. J. 1982. Know thy memory: The use of questionnaires to assess and study memory. *Psychological Bulletin 92,* 2, 434–452.

Huttenlocher, J. and Prohaska, V. 1997. *Memory for Everyday and Emotional Events*. Lawrence Erlbaum Associates, Chapter Reconstructing the Times of Past Events, 165–179.

Jones, W., Bruce, H., Foxley, A., and Munat, C. 2005. The universal labeler: Plan the project and let your information follow. In *Proceedings 68th Annual Meeting of the American Society for Information Science and Technology (ASIST) 42, Charlotte (US)*, Grove and Andrew, Eds.

Jones, W. P. and Dumais, S. T. 1986. The spatial metaphor for user interfaces: experimental tests of reference by location versus name. *ACM Trans. Inf. Syst. 4,* 1, 42–63.

Kalnikaité, V. and Whittaker, S. 2007. Software or wetware?: discovering when and why people use digital prosthetic memory. In *CHI '07: Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM, New York, NY, USA, 71–80.

Kaptelinin, V. 2003. Umea: translating interaction histories into project contexts. In *CHI '03: Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM Press, New York, NY, USA, 353–360.

Kelly, D. and Teevan, J. 2007. *Personal Information Management*. Seattle: University of Washington Press., Chapter Understanding what works: Evaluating personal information management tools, 190–204.

Kerr, N. H. 1983. The role of vision in "visual imagery" experiments: Evidence from the congenitally blind. *Journal of Experimental Psychology 112*, 265–277.

Kosslyn, S. M. 1981. The medium and the message in mental imagery: a theory. *Psychological Review 88*, 46–66.

Kwasnik, B. H. 1989. How a personal document's intended use or purpose affects its classification in an office. *SIGIR Forum 23,* SI, 207–210.

Lansdale, M. 1988. The psychology of personal information management. *Appl Ergon 19,* 1, 55–66.

Lansdale, M. and Edmonds, E. 1992. Using memory for events in the design of personal filing systems. *International Journal of Man-Machine Studies 36,* 1, 97126.

Lansdale, M. W. and Simpson, M. 1990. A comparison of words and icons as external memory aids in an information retrieval task. *Behaviour & Information Technology 9*, 11–131.

Mackay, W. E. 1988. More than just a communication system: diversity in the use of electronic mail. In *CSCW '88: Proceedings of the 1988 ACM conference on Computer-supported cooperative work*. ACM, New York, NY, USA, 344–353.

Malone, T. W. 1983. How do people organize their desks?: Implications for the design of office information systems. *ACM Trans. Inf. Syst. 1,* 1, 99–112.

Mander, R., Salomon, G., and Wong, Y. Y. 1992. A pile metaphor for supporting casual organization of information. In *CHI '92: Proceedings of the SIGCHI conference on Human factors in computing systems*. ACM Press, New York, NY, USA, 627–634.

McCullagh, P. and Nelder, J. A. 1989. *Generalized Linear Models*, second ed. Chapman & Hall/CRC.

Ringel, M., Cutrell, E., Dumais, S., and Horvitz, E. 2003. Milestones in time: The value of landmarks in retrieving information from personal stores. In *Proc. INTERACT 2003*. 184–191.

Robertson, G., Czerwinski, M., Larson, K., Robbins, D. C., Thiel, D., and van Dantzich, M. 1998. Data mountain: using spatial memory for document management. In *UIST '98: Proceedings of the 11th annual ACM symposium on User interface software and technology*. ACM Press, New York, NY, USA, 153–162.

Rogers, T. B., Kuipers, N. A., and Kirker, W. 1977. Self-reference and the encoding of personal information. *Journal of Personality and Social Psychology 35*, 677–688.

Rubin, D. 1977. Very long-memory for prose and verse. *Journal of Verbal Learning and Verbal Behavior 16*, 611–621.

Rubin, D. and Wenzel, A. 1996. One hundred years of forgetting: A quantitative description of retention. *Psychological Bulletin 103*, 734–760.

Sachs, J. S. 1967. Recognition memory for syntactic and semantic aspects of connected discourse. *Percept. Psychophys 2*, 437–42.

Schraefel, M. C., Smith, D. A., Owens, A., Russell, A., Harris, C., and Wilson, M. L. 2005. The evolving mspace platform: leveraging the semantic web on the trail of the memex. In *Proceedings of Hypertext, 2005, Salzburg*.

Sellen, A. J. and Harper, R. H. R. 2003. *The Myth of the Paperless Office*. MIT Press, Cambridge, MA, USA.

Thompson, C. P., Skowronski, J. J., Larsen, S. F., and Betz, A. 1996. *Autobiographical Memory: Remembering What and Remembering When*. Lawrence Erlbaum Assoc Inc.

Tulving, E. 1983. *Elements of Episodic Memory*. Oxford University Press, New York.

Whittaker, S., Jones, Q., and Terveen, L. 2002. Contact management: Identifying contacts to support long term communication. In *Proceedings of CSCW 2002 Conference on Computer Supported Cooperative Work*. New York: ACM Press., 216–225.

Whittaker, S. and Sidner, C. 1996. Email overload: exploring personal information management of email. In *CHI '96: Proceedings of the SIGCHI conference on Human factors in computing systems*, M. J. Tauber, Ed. ACM Press, New York, NY, USA, 276–283.

Wittrock, M. and Carter, J. 1975. Generative processing of hierarchically organized words. *American Journal of Psychology 88*, 489–501.

Yang, Y. 1994. Expert network: effective and efficient learning from human decisions in text categorization and retrieval. In *SIGIR '94: Proceedings of the 17th annual international ACM SIGIR conference on Research and development in information retrieval*, W. B. Croft and C. J. van Rijsbergen, Eds. Springer-Verlag New York, Inc., New York, NY, USA, 13–22.

Yee, K., Swearingen, K., Li, K., and Hearst, M. 2003. Faceted metadata for image search and browsing. In *CHI '03: Proceedings of the SIGCHI conference on Human factors in computing systems*, G. Cockton and P. Korhonen, Eds. ACM Press, New York, NY, USA, 401–408.

...