A Study of Remembered Context for Information Access from Personal Digital Archives

Liadh Kelly
Centre for Digital Video Processing
School of Computing, Dublin City University
Dublin 9, Ireland
lkelly@computing.dcu.ie

Yi Chen
Centre for Digital Video Processing
School of Computing, Dublin City University
Dublin 9, Ireland
ychen@computing.dcu.ie

Marguerite Fuller
Centre for Digital Video Processing
School of Computing, Dublin City University
Dublin 9, Ireland
mfuller@computing.dcu.ie

Gareth J.F. Jones
Centre for Digital Video Processing
School of Computing, Dublin City University
Dublin 9, Ireland
gjones@computing.dcu.ie

ABSTRACT

Retrieval from personal archives (or Human Digital Memories (HDMs)) is set to become a significant challenge in information retrieval (IR) research. These archives are unique in that the items in them are personal to the owner and as such the owner may have personal memories associated with the items. It is recognized that the harnessing of an individual's memories about HDM items can be used as context data (such as user location at the time of item access) to aid retrieval. We present a pilot study, using one subject's HDM, of remembered context data and its utility in retrieval. Our results explore the types of context data best remembered for different item types and categories over time and show that context appears to become a more important factor in effective HDM IR over time as the subject's recall of contents declines.

Topic Area.
Case studies, field experiments, simulations, etc. of context-sensitive information seeking & retrieval.

1. INTRODUCTION

Recent years have seen individuals storing increasing amounts of personal information in digital format. We have now reached the point where many of a person's personal life experiences can be stored digitally - everything from items read, written, or downloaded; to footage from life experiences, e.g. photographs taken, videos seen, music heard, details of places visited, details of people met, etc, can all be captured using devices such as computers, mobile phones, cameras, video recorders, audio recorders, GPS technology, sensor technology etc.

In order to make these new archives more valuable than "a vast unintelligible mountain of information", in which locating the required information becomes synonymous with locating "a needle in a haystack", steps need to be taken to develop effective means of retrieval from them. While existing information retrieval (IR) techniques are good at locating relevant items from traditional data archives in response to users' queries, Human Digital Memories (HDMs) form a new type of archive requiring new retrieval techniques. This domain is fundamentally distinct from traditional IR domains in that: there is the potential for a large percentage of noisy data in these archives; many items in the archive may be very similar, repeatedly covering the same topic; items in HDMs are personal to the individual and the individual will have personal memories about items related to such things as time and place of item creation and subsequent access. These factors combined lead to the requirement of new retrieval techniques specific to this domain.

Existing studies into recalled attributes of computer files have found that people usually do not have very good memory of individual items of information, rather they have better memory of personal experiences at the time of using/creating the item/file (e.g. remembering creating a document on a sunny day) [5]. This kind of memory is called episodic or autobiographical memory, which refers to memory of experiences involving 'ones self'; while memory of individual items of information is usually referred to as semantic memory, which corresponds to facts or knowledge without autobiographical context [18]. Studies on source memory also showed that people have more accurate and detailed memory of context such as relative position of the item than the item itself [7]. For example, an individual may not remember the name of a photo, but may remember that it was taken in Paris, with a friend, on a sunny day, and that it was in the afternoon. Thus, if files were tagged with episodic context information, such as location and time of the file/information processing, it would be possible for an individual to query based on recalled episodic memories.

We believe memory relating to the context of information capture, and potentially the context of earlier re-finding of items, can play a significant role in IR for the HDM domain. Within our work we are exploring which types of (episodic) context data are best remembered by individuals and the

utility of this context data for search within HDMs. We hope to use the context information that best matches both criteria, to tag the files/information, to help facilitate retrieval. Of particular interest is establishing the types of context data which are most useful for this task, and how this context data can be used to link and annotate items to increase retrieval performance from HDMs. A pilot study we conducted to explore these issues is presented in this paper.

This paper is structured as follows: Section 2 reviews existing work in systems for searching HDMs and summarises cognitive studies relating to this field; Section 3 describes our pilot study; and finally Section 4 concludes with directions for further research and experimentation.

2. RELATED WORK

A number of existing studies have explored the use of people's recalled memory of past interactions with items as a means to relocate them. Research such as [1], [5] and [17] discovered many attributes individuals recall about files such as location of file, actions performed on file, daylight status, weather and local time which they use as context data to retrieve them.

In other research Elsweiler et al [5] used their memory study findings to develop a photo browser which exploits people's remembering mechanisms. There are many other examples of the use of context in simple ways in existing work. The Microsoft MyLifeBits [8] and Stuff I've Seen (SIS) [4] systems for example make use of context data, such as date and file type, to enable an individual search for personal files.

We postulate that individuals may recall different types of context data for different file types, and the depth of their recollection may be dependant on the type of item being retrieved. Different context may be important for different data types, and different context may be recalled in different situations. In our work we are examining if there is a relationship between recalled context and the usefulness of specific context features for retrieval.

2.1 Memory and Remembering

Psychology studies on long term memory have shown that well remembered information usually has deeper processing at the time of encoding [3][11]. However, due to an individual's limited cognitive processing resources [13], the more distinctive and closely related to previous knowledge or the present situation the information is, the easier it may be to link to it, and the more elaborately it will be encoded. This means that it can be easier to remember items which are related to oneself or with concepts that are familiar. One example is the generation effect [10], which states that things one thought of or did oneself tend to be better remembered, because this kind of information will be linked with one's experience of creating it.

Apart from the strength of encoding, successful recall also depends on the ease with which an item can be retrieved, which relates to how effective the cues presented during retrieval are [16]. One influential hypothesis called 'encoding specificity', which was postulated by [19] and supported by many classic studies [9] and recent neuroimaging findings [15], states that the effectiveness of retrieval depends on how similar the conditions of retrieval are with that of encoding. While an individual is searching for a file in their HDM, the target file acts as the cue for the searcher to retrieve related information regarding this file from their memory, this related information forms the search query. According to the "encoding specificity" view, information from a similar cognitive category (e.g. visual spatial, verbal acoustic, semantic) to the file is more likely to be remembered.

We postulate that users remember different types and levels of context data depending on the type of file they are accessing, and that if this is harnessed it will be possible to create more efficient and effective retrieval systems. Towards this we conducted a preliminary study which examines the types of context data best remembered in different situations, and the effects of these findings on HDM item retrieval.

3. PILOT USER STUDY

In this section we describe a pilot study to test the types of context data recalled by an individual and the utility of this context data in HDM retrieval. The results of this study suggest the types of context data an individual is most likely to recall in the long-term and that this context data is beneficial in HDM retrieval. We begin by describing the user study conducted and then analyse the results obtained.

3.1 Data Collection

A small experimental HDM was created to allow us examine the context data recalled (and utility of this context data) by the HDM owner directly following data collection and after a 6 month interval. The Mac OS X computer activity of one subject was recorded over a period of six weeks using Slife[1]. Slife records the event of a window being brought to the foreground. In total roughly 10,000 events were recorded over the 6 week period. For each event it records: the textual content inside the window (e.g. the text of an email, web page or document being written), time and date accessed, the duration the window was open for, events linked to if applicable, type of application (e.g. web, chat), document, source (e.g. Microsoft Word), and window title. The following additional context data was also created for events:

- Using time and date, information functions were written to determine, the hour, minute, second, season and period of the day in which the event took place e.g. morning, afternoon, evening, night.
- Linking of related events. This involved slicing each day into events consisting of bursts of activity, where a slice partition occurred if the time between events exceeded 30 minutes. This provided the following additional context data for event objects: types of events in the slice (i.e. surrounding event type), sources of events in the slice (i.e. surrounding event source) and content of events in the slice (i.e. surrounding event content).
- Geographic location, e.g. office, lab, etc was captured by prompting the subject to enter their current location on their computer at startup and at varying intervals during computer use. All events in a given partition/slice were then annotated with the location data occurring at the median time of the partition.

[1]http://www.slifelabs.com

- Weather data. Weather history was obtained from [12], and parsed to extract the weather history for each hour of a day. In a similar fashion to location data, the weather data occurring at the median time of a partition was annotated to all events in that partition/slice.

Lucene[2], an open source search engine, was used to index events (i.e. items and their associated context data). Using Lucene the context data associated with events was indexed into different fields (e.g. location field, weather field etc). The StandardAnalyzer built into Lucene was used to index the content of events. This tokenizes the content based on a sophisticated grammar that recognises email address, acronyms, alphanumeric and more; converts lowercase, and removes stopwords. Table 1 provides a summary of the complete set of context data associated with events. Full details of the collection and preprocessing of this data set are described in [6].

Event ID
Event content
Title
Minute, Hour, Date, Month, Year
Weekday e.g. Mon, Tues
Season e.g. summer, winter
Source e.g. Word, Firefox
Type e.g. document, chat, Web
Location e.g. college, kitchen
Weather e.g. showers, cloudy
Surrounding Events Types
Surrounding Events Sources
Surrounding Events Content

Table 1: Summary of the complete set of data associated with events

3.2 Test Case Generation

On completion of the computer activity recording process a set of 30 content re-finding test case scenarios were created from the participant's memory without looking at the data set. These test cases were drawn from events that covered the time span of the data collection from middle of July 2007 to the end of August 2007. To generate the test cases the participant identified the key events that occurred during the six week data collection period, these included: a friend's birthday, several meetings in the office, and dinner in a restaurant. Following this, the participant was required to recall the activities performed on their computer around or close to these key events and other context data. The following is an example of a typical recall: *"I was in work on a Friday in July, I remember thinking I couldn't find my umbrella that morning, and I hoped it wouldn't rain. I decided to meet friends in the Golden Lion restaurant for dinner after work. It was fairly late when I went to get the bus home, however, it wasn't the last bus. On the way to the bus stop, I remember thinking how lucky I was that it wasn't raining, as I didn't have my umbrella with me. On returning home, I logged onto the computer in my office. My friend, Sarah, was on-line, I remember telling her about the tasty curry I had for dinner".* The recorded data for each test case then consisted of a remembered scenario and computer file (e.g. email, word document, web page) accessed around or close to the scenario and other recalled related context data.

If HDMs are to be recorded and accessed over an extended period it is important that users are able to reliably retrieve content recorded in the distant past. It is clear that a user is likely to remember a significant amount of context data soon after an event occurred, however with time memory fades and it is anticipated that less will be remembered after a substantial delay after the event occurred, as discussed in Section 2.1. To explore this effect and its potential impact on re-finding effectiveness, we conducted a further study on the subject 6 months after the initial test case generation process. Results of this later study are described in Section 3.5.

3.3 Query Types

After establishing the 30 test case scenarios, the remembered data (we refer to this as 'initially remembered data') from the scenarios was converted into queries. Remembered data 6 months later, relating to the required computer files from the 30 test case scenarios, was also converted into queries. The remembered data 6 months later was established using the technique described in Section **??**.

Different types of queries were constructed to assess the usefulness of the various types of context data on their own and also in combination. Eleven types of queries relating to each re-finding test case were constructed for the initially recalled data and for that remembered 6 months later.

- Query One: Content only
- Query Two: Context only, this incorporated, the following fields: title, source, type, location, weather, year, month, day, hour, minute, weekday, period, surrounding event types, surrounding event source, surrounding event content, season.
- Query Three: Combination of Content and Context.
- Query Four: Combination of Content and Time, i.e. hour, minute, weekday, day, month, year, season and period of the day.
- Query Five: Combination of Content and Weather.
- Query Six: Combination of Content and Location.
- Query Seven: Combination of Content and Type and Source.
- Query Eight: Combination of Content and Surrounding Type, Content and Source.
- Query Nine: Combination of Content and Day.
- Query Ten: Combination of Content and Period of day.
- Query Eleven: Combination of Content and Date.

Query type one represents the current standard approach for retrieval using search engines. Results generated from content only queries were used as the benchmark. Query types three to eleven are straightforward concatenations of the content data from query one with various types of the context data from query two.

[2]http://lucene.apache.org/java/docs/

3.4 Querying

Re-finding specific files or items is referred to as a known-item search. Our task was thus a known-item search to retrieve the correct file for a given remembered test case scenario. To investigate the usefulness of context data in the retrieval process the 11 sets of 27 query types[3], using the context data remembered immediately following data collection and that recalled 6 months later, were entered into Lucene and the rank of the target document in the result set was noted. Examples of queries, based on the test case example presented in Section 3.2, include:

- Query One: content: (+ curry + golden + lion)
- Query Two: type: chat source: adium year: 2007 month: July period: night weekday: Friday season: summer location: office weather: cloudy surType: web surSource: firefox
- Query Three: content: (+ curry + golden + lion) type: chat source: adium year: 2007 month: July period: night weekday: Friday season: summer location: office weather: cloudy surType: web surSource: firefox
- Query Four: content: (+ curry + golden + lion) year: 2007 month: July period: night weekday: Friday season: summer

3.5 Study 1 - Remembered Content and Context

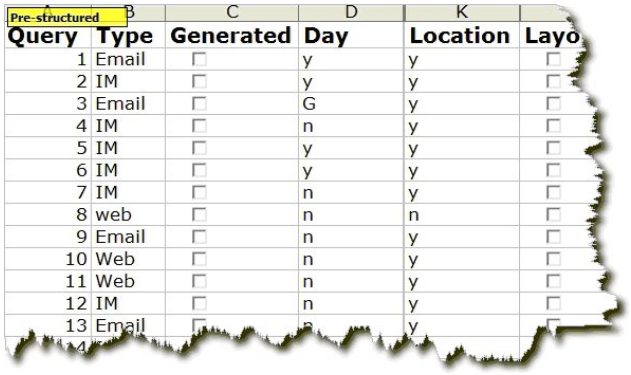
3.5.1 Recall of search test cases after 6 months

As discussed previously, memories associated with items in an HDM will often fade over time after the event. In order to begin to explore this effect as part of our pilot study we conducted a further study with our test participant 6 months after the HDM data collection.

A simple desktop application was used for a free recall test. This was used in preference to an oral report strategy in consideration of user privacy, and since it is more natural, people generally type queries to search on the computer, as opposed to orally reporting their queries.

The participant was given a demo and then, to get them accustomed to the application and to estimate the time that would be required to complete the formal task, performed five practice tasks with information given by the system developer which they were asked to enter on the experimental platform. For the formal task, the participant was asked to free recall targets which she searched for in her previous study 6 months ago, and all the relevant information about them including both content and context (this was not confined to only what was used previously). Abbreviations were allowed to reduce the time needed to enter long words/phrases or sentences. Since the participant did not recall all the 'original test cases', the free recall task was followed by a cued recall session to aid recollection of further 'original test cases'. This consisted of presenting the participant with a list of titles for the unrecalled 'original test cases'. With this list of titles as cues, the participant recalled further information on these test cases. This was followed by a post-test interview to verify if the subject had in fact recalled the episodic facts for surrounding events, location and weather, as opposed to simply recalling the keywords used in the 'original' retrieval experiments conducted 6 months earlier.

[3]Due to technical difficulties only 27 of the 30 test cases were available for querying.



Pre-structured

Query	Type	Generated	Day	Location	Layo
1	Email	☐	y	y	☐
2	IM	☐	y	y	☐
3	Email	☐	G	y	☐
4	IM	☐	n	y	☐
5	IM	☐	y	y	☐
6	IM	☐	y	y	☐
7	IM	☐	n	y	☐
8	web	☐	n	n	☐
9	Email	☐	n	y	☐
10	Web	☐	n	y	☐
11	Web	☐	n	y	☐
12	IM	☐	n	y	☐
13	Email	☐	n	y	☐

Figure 1: Pre-structured result sheet ***note: n=not recalled; y= correctly recalled; g=guessed and partly correct; rate of content remember ranges from 0=not at all to 4 =very well including the title and clear details. in the field of content is the number of points recalled;***

In order to establish what content and context information the participant recalled, did not recall, partially recalled and guessed for each test case, the participant compared the recalled content and context information with the original records generated 6 months earlier. The results of this were entered into a pre-structure Excel spreadsheet (see Figure 1).

An in-depth interview was then carried out with the participant based to the results in the spreadsheet. The interview tried to clarify whether the recalled details were from the memory of the previous tests or from that of the original experiences about using the files or applications.

3.5.2 Data Analysis

SPSS 14.0 was used to analyse the collected data. For each field in the pre-structured Excel spreadsheet, all the correctly recalled data was replaced with 1, missed or incorrectly recalled with 0, and those that the subject marked as guessed (correct or partly correct, e.g. the day should be Tuesday, but "either Tuesday or Wednesday" was recorded) with 0.5. The percentage of correct recall, and guessed (if applicable) was calculated. The 30 queries were categorized by:

1. Source type (Instant Messaging (IM), Email, etc.),
2. Self-generated vs Passively Presented; the context information was also grouped by Visual/Spatial (perceptual) vs. Textual (semantic, conceptual).

The overall recall score from the two sections (free recall vs. cued recall with file title) did not show consistent difference (see Figure 2). It is possible that although cued recall could trigger more memory than free recall, the participant was more familiar with the queries that she recalled in the first section. Across both sections, the subject had particularly good memory for the file types. The subject's location, month and period of the day were also well remembered.

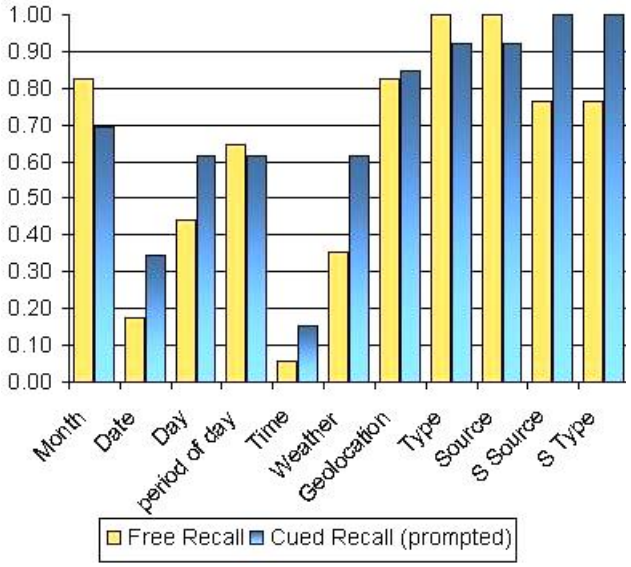


Figure 2: Context Recall Results from free recall and cued recall. **note: S source= Sorrounding events source; S type= surrounding applications' type*

On average the subject recalled less content points for each query than she previously used. Their average self rating of "how well the content is remembered" was 2.36 out of 4 (0-4), where the score immediately after data collection would have been nearer to 4. According to the interview, this was largely due to a considerable portion of well recalled details of associated with visual or non narrative elements, but very limited recall of textual information.

Memory of Perceptual vs Conceptual information.

The average recall percentage of more perceptual type of memory (e.g. location, weather, period of the day) is overall better than the verbal conceptual(textual) (e.g. month, exact date/day, time) type (Mean Difference (MD) =0.22, t (29) =3.25, p<.005), the gap is even greater when the searching target is self generated files (MD=0.386).

Memory of information for different target types.

The memory of each type of information (both content and context) does not differ significantly for the different file types. However, there is a difference for self generated information types (e.g. IM, some documents) and types presented by others (e.g. Web, iTunes).

Self generated vs Passively presented.

The difference in recall performance between self generated and passively presented files is significant for the overall recall score of context, especially of location, and surrounding application. However, for date and time, weather, and the content itself, the generation effect did not show much influence. This is not consistent with [10], according to which, recall of content should be particularly affected by whether it is self generated or passively presented. Our results indicate that perceptual memory might be more sensitive to personal experience involving oneself, that is, the more personal involvement required, the better perceptual context tends to be remembered.

3.5.3 Discussion

We tested the subject's memory on 30 queries which she created 6 months ago. We found that textual content was not well recalled, while some context information showed promising recollection. According to the interview, the subject relied on her episodic memory of what was happening during days associated with sought for items, to recall most of the context information, which thus seems to be much better retained than content of the files. The effect of self generated lower level or perceptual level information, such as visual memory, while better remembered, seems to rely more on episodic memory, which is also claimed to be largely perceptual [2], than does conceptual memory. In short, the results support our hypothesis that the types of remembered information differ according to different targets and personal behaviour towards it.

3.6 Study 2 - Querying Results

For each query, the rank of the target document in the list returned by Lucene was noted. The Mean Rank and the Mean Reciprocal Rank were then calculated for each query set. The Mean Rank as its name suggests, is the average rank of the target documents. The reciprocal rank of a query response is the multiplicative inverse of the rank of the correct answer. The mean reciprocal rank is recommended as a measure of evaluation for known item search as it is not severely influenced by target documents retrieved at low ranks [14].

The results we obtained on an initial study of this data [6], using the context data recalled directly proceeding data capture, showed the combination of content and context data to perform very well.

This combination performed considerably better than content only with an increase of 0.21 in the results, using the 27 available test cases (see Figure 2). Recalled context and content 6 months later shows an even greater improvement over content only (increase of 0.3, (see Figure 3)). The surprisingly low mean rank of 88 for content and context combined can be explained by one of the queries for a web page having a rank of 2,181. Consistent with the original results, recalled content and time, and content and location after a 6 month interval performed well.

Comparing the mean reciprocal rank results, advantage was found for combining remembered content with context as opposed to using remembered content only for both the initially recalled and recalled 6 months later data. Interestingly, for the recalled data 6 months later this advantage is *significantly* greater than it is for the initially recalled data(t(26) =2.53, p<0.05). These results suggest support for the use of context data in retrieval, and that over time as individual's memory of content data associated with HDM items fades the use of context data in retrieval becomes more important.

Overall, source, surrounding events and time associated with test cases were well remembered. Each of these items combined in isolation with recalled content dramatically improved the rank of the target item in the results (see Figure 3), which suggests that a retrieval system which allows users to search based on these types of context data would be very beneficial.

Query Type	Mean Rank	Mean Reciprocal Rank
Content only	9.07	0.49
Context only	39.93	0.26
Content & Context	2.85	0.70
Content & Time	7.15	0.59
Content & Weather	13.19	0.37
Content & Location	5.22	0.6
Content & Type and Source	4.81	0.49
Content & Surrounding Type, Content and Source	12.15	0.39
Content & Day	4.41	0.69
Content & Period	4.19	0.71
Content & Date	8.64	0.74

Table 2: Experimental Results - initially recalled data

Query Type	Mean Rank	Mean Reciprocal Rank
Content only	213.11	0.33
Context only	319	0.28
Content & Context	88.12	0.63
Content & Time	8.13	0.67
Content & Weather	15.18	0.38
Content & Location	9.96	0.48
Content & Type and Source	7.64	0.41
Content & Surrounding Type, Content and Source	14.73	0.33
Content & Day	203.74	0.54
Content & Period	106.69	0.56
Content & Date	9.73	0.68

Table 3: Experimental Results - recalled data after 6 month interval (free recall and cued recall combined)

The granularity and features of remembered time, decreased over the 6 month period. Despite this, recalled content combined with time yielded the best results. We were interested in establishing which factors of time in particular might be useful to include in a HDM search system. Unsurprisingly exact date combined with content showed the best results for the cases where date was recalled. Both content + day (e.g. Monday, Tuesday), and content + period of day (e.g. morning, evening) performed quite well. The low mean ranks here are largely attributable to the poor retrieval performance of a few test cases - omitting these test cases would have resulted in significantly better retrieval performance - suggesting that further investigation is necessary to determine the utility of day and period of day in retrieval from HDMs collected over extended periods or accessed a long time after collection.

Of particular interest, in calculating these results the test subject was unsure of the exact date or day for some test cases, and thus performed queries of the form '*date1* or *date2*' and '*day1* or *day2*' in these instances. This imprecision in recollection highlights the need for search options of a lower granularity, e.g. weekday and weekend, as opposed to exact days.

4. CONCLUSIONS AND FUTURE WORK

In this paper we presented a pilot study which investigated the types of HDM data individuals are more likely to recall in the longer term and the utility of this data in HDM retrieval. In our pilot study using one person's personal data recorded over a period of 6 weeks, the recall results indicate that while much of the narrative/textual content failed to be recalled, context information (such as the location which is closely related to personal experiences at the time of encoding) and file types (which easily tend to be triggered by the queries/targets themselves) were well remembered over the 6 month period. These findings are largely consistent with the psychology studies. The retrieval results obtained are also promising and show that over the longer term recalled context data can be used to improve content only retrieval performance in the HDM domain.

Of particular interest, while no significant reduction in the amount of content recalled 6 months later was noted due to the limits of the test, a dramatic decrease in the performance of content only query results using the content recalled 6 months later was observed. This occurred despite the fact that the test subject was an expert searcher, although the importance or otherwise of this factor would need to be investigated in a further study. Our findings suggest that over time recalled content is of a lower quality and a considerable amount of the important key words are likely to be forgotten. This provides further support for the notion that context is important for retrieval as users more readily remember context associated with files over the longer term than the actual keywords contained within the files.

While these results are promising, it is acknowledged that the collection was limited in size - it only spanned six weeks, HDMs have the potential to span many years, a lifetime. Searching involved a few thousands documents compared to possible millions. Additionally, the results are specific to one person. Investigation using more participants, over a longer time frame is planned to further test our conclusions.

5. ACKNOWLEDGMENTS

This work is funded by a grant under the Science Foundation Ireland Research Frontiers Programme 2006.

6. REFERENCES

[1] T. Blanc-Brude and D. Scapi. What do people recall about their Documents? Implications for Desktop Search Tools. In *IUI'07*. Honolulu, Hawaii, USA, January 2007.

[2] M. Conway, J. Singer, and A. Tagini. The self and autobiographical memory: Correspondence and coherence. *Social cognition*, 22:491, 2004.

[3] F. Craik and R. Lockhart. Levels of processing: A framework for memory research. *Information Processing and Management*, 11:671–684, 1972.

[4] S. Dumais, E. Cutrell, J. Cadiz, G. Jancke, R. Sarin, and D. Robbins. Stuff I've seen: a system for personal information retrieval and re-use. In *SIGIR '03: The*

26th annual international ACM SIGIR conference on Research and development in informaion retrieval, pages 72–79, New York, NY, USA, July 2003. Toronto, Canada, ACM Press.

[5] D. Elsweiler, I. Ruthven, and C. Jones. Towards Memory Supporting Personal Information Management Tools. *Journal of the American Society for Information Science and Technology*, 2007.

[6] M. Fuller, L. Kelly, and G. Jones. Applying Contextual Memory Cues for Retrieval from Personal Information Archives. In *PIM 2008 - Proceedings of Personal Information Management, Workshop at CHI 2008*. Florence, Italy, 2008.

[7] J. Gardiner. Episodic memory and autonoetic consciousness: a first-person approach. *Philosophical Transactions: Biological Sciences*, 356(1413):1351–1361, 2001.

[8] J. Gemmell, G. Bell, and R. Lueder. MyLifeBits: A Personal Database for Everything. *Communications of the ACM. Personal Information Management*, 49(1):88–95, January 2006.

[9] D. Godden and A. Baddeley. Context-dependent memory in two natural environments: on land and underwater. *British Journal of Psychology*, 66(3):325–331, 1975.

[10] R. Greene. *Human memory: Paradigms and paradoxes*. Hillsdale, NJ, Erlbaum, 1992.

[11] V. Gregg and J. Gardiner. Recognition Memory and Awareness: A Large Effect of Study - Test Modalities on "Know" Responses Following a Highly Perceptual Orienting Task. *European Journal of Cognitive Psychology*, 6:131, 1994.

[12] Ireland Weather. Hourly weather history for DUBLIN AIRPORT, Ireland. http://www.freemeteo.com.

[13] D. Kahneman. *Attention and Effort*. Englewood Cliffs, NJ: Prentice-Hall, 1973.

[14] P. Kantor and E. Voorhees. The TREC-5 Confusion Track: Comparing Retrieval Methods for Scanned Text. *Inf. Retr. (IR)*, 2(2/3):165–176, 2000.

[15] H. Park and M. Rugg. The Relationship between Study Processing and the Effects of Cue Congruency at Retrieval: fMRI Support for Transfer Appropriate Processing. *Cereb Cortex*, 18(4):868–875, 2008.

[16] R.A. Bjork and E.L. Bjork. *A new theory of disuse and an old theory of stimulus fluctuation*, pages 35–67. Erlbaum, Hillsdale, 1992.

[17] J. Teevan, C. Alvarado, M. Ackerman, and D. Karger. The Perfect Search Engine Is Not Enough: A Study of Orienteering Behavior in Directed Search. In *CHI 2004: Conference companion on Human factors in computing systems*, pages 415–422. Vienna, Austria, April 2004.

[18] E. Tulving and W. Donaldson. *Episodic and semantic memory*, pages 381–403. NewYork: Academic Press, 1972.

[19] E. Tulving and D. Thomson. Encoding specificity and retrieval processes in episodic memory. *Psychological Review*, 80:352–373, 1973.