Exploring Memory Cues to Aid Information Retrieval from Personal LifeLog Archives

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

ABSTRACT

The expansion of personal information archives and the emerging field of Personal Lifelogs (PLs) are creating new challenges for information retrieval (IR). While studies have demonstrated the difficulties of IR for these massive data collection [1], we should also think about how we can opportunities and benefits from integrating these data sources as a component of "digital memories" , considering their rich connections with the users' memory. We observed that most existing approaches to personal archive IR are mostly technology-driven. Although in recent years studies in Personal Information management (PIM) have claimed to make use of the human memory features, and many works have been reported as investigating well-remembered features of computer files (documents, email, photos). Yet, these explorations are usually confined to the attributes or feature that current computer file systems or technology have provided.

I believe that there are important and potentially useful data attributes that these studies have ignored. In addition, current personal search interfaces provide searching options based on what is available in the system, e.g. require users to fill in the calendar date, regardless of the fact that people actually don't often encode 'time' in such a way. My PhD project aims to explore what users actually tend to recall in different personal achieve information seeking tasks, how to present searching options to cater for the right type or format of information that users can recall, and how to exploit this information in an IR system for personal lifelog archives.

In this paper, I discuss the limits and advantages of some related work, and present my current and proposed study, with an outlook of an interface that I plan to develop to explore my proposals.

Keywords

Searching interface, Research methods, human memory, personal information management

1. INTRODUCTION

Computing activities play an increasingly important part in our daily lives, no longer confined to electronic documents in the office, but contributing to our entertainment at home, e.g. music, YouTube, our communication with others, e.g. emails, chat with instant message (IM), sharing experiences or diaries via blogs, and our plans and schedules. Beside, more and more aspects of our real life can be captured by digital devices to store and use on the computers. The prevalence of cameras, voice recorders, location devices, in particular, the introduction of automatic capturing/recording devices such as the Microsoft Sensecam means that rich archives of moments from our lives can be captured and stored. Recording a life experience of what one saw and heard is what referred to as *life logging*. A very detailed life log could be seen as something like a static digital copy of the human memory as outlined by Sparck Jones [2]. She suggested that such an archive can be used as a 'deposit' where currently unimportant information can be stored, so that the human memory can work more efficiently with less but more useful data, and come back to retrieve these deposited items in the future when they are needed. Or it can augment human memory, reminding people of what may be of potential interest, but they might have forgotten. This data might be selectively presented to others to retell details from one's life stories. Such data collections are currently being built up by increasing numbers of researchers, most visibly by pioneers such as Gordon Bell with the support of archiving tools such as Microsoft MyLifeBits [3] with the Miscrsoft SenseCam, an wearable camera for automatic capturing real life experiences proactively [4].

Yet, these Personal Lifelogs (PLs) archives can be of little use if useful information in them cannot be identified and retrieved. There have been many systems engaging in personal information management (PIM) for various types of data, e.g. the management of emails, a library to manage music or video clips, or photo albums to manage photos. However, it is difficult to perform search tasks where multiple types of data are involved, or where the data type is not clear. Beyond this issue, management tools do not exist for many types of data, and where they do, the efficiency of these searching tools can be poor. On the other hand, 'all-in-one' IR tools such as windows desktop search and Google Desktop, while indexing most items within a computer, only enable searching based on keywords from textual content or files name, or last accessed or modifies date, etc. However, they are unlikely to be the things that users tend to remember.

In the following section we will show some example of related work which explores or uses well remembered data attributes, and discuss the limits of these studies. In Section 3, I will present my proposed study, and in Section 4 will outline the scope of my planned experimental system.

2. RELATED WORK

2.1 Searching options to include

There have been various studies suggesting the use of well remembered attributes.

The Stuff I've Seen (SIS) system removed barriers between item types and enabled the search of data from files to WebPages,

email, etc., and enabled the search by comparatively better remembered memory cues attributes, such as authors [15].

Blanc-Brude tested memory of a document's attributes on 14 participants [5]. Their conclusion mainly comes from a cued recall test, in which they used the names of attributes taken from major PIM literature, as cues to assist subjects' recall. Thus their main findings may have been limited by the types of attributes used. Although their experiments included a free recall phase, it was also to some extent guided the subjects by the instructions to recall the listed features of the documents.

There are also some similar studies exploring remembered features of more real life related types of data. For example, [6] tested subjects' memory with specified questions about physical features (e.g. where, who) on meeting videos; [7] instructed subjects to rate their memory of a group of pre-listed features of their free recall of three photographs. A limitation of [7] is that the firstly/priority free recalled photographs are usually well-remembered ones, which means a sample bias that makes the result not applicable for a wide range of less well remembered photographs in one's collection. In fact, all of the above studies [5, 7] only tested subjects' memory on known items, meaning that their results of recall performance on attributes can only be applied to the situation where the user knows which specific file to look for and searches for that known item. However, people actually don't remember the existence of many items, so it is not possible for them to search for such unknown items as targets, let alone recalling useful corresponding search features. Information seeking studies suggest that behaviour towards these types of data is usually undirected, e.g. browsing for what is available [8]. This implies that some strategies of embedding browsing in a user interface could enable users to access potentially interesting information which they have forgotten about.

Studies on information re-finding behaviour, e.g. [9], also indicate that factors such as task type and elapsed time may influence a user's the performance on the task. For this reason, [10] emphasized the importance of non-intrusive observation of a user's behaviours in a natural context, and carried out a web-based diary study in which the subjects took notes including the reason for searching and the target (what the subject wished to find) after their natural searching tasks on emails and web content. They concluded that there were three types of frequent searching target: looking up some information, searching for a single known file, or getting information from multiple items. Yet, their later study on memory of emails was also confined to the standard attributes that emails possess. Other examples of research or application of this idea include [11, 12], which also successfully integrated context (metadata) from the digital item themselves and from real world. But most of these systems provided options that stemmed from what the current technology provides, although they selected better remembered features from these existing options.

I believe that there are potentially useful and feasible (both well-remembered and possible for current techniques to realize) types of information that have been omitted by the above approaches. We aim to jump out of the enclosure of the currently existed features, and give users the freedom to include what they tend to recall, and then we developers can judge based on the feasibility of attribute capture techniques to provide these options.

In one of our pilot studies[13], the subject collected personal life log data over a 6 week period, and then generated 30 information seeking scenarios from this period in a semi-natural way. She took notes of what happened around this time from her free recall, e.g. a meeting with friends. Although we could not draw any statistically valid conclusions of well remembered attribute types from a single subjects' data, the study did suggest a possible inclination of users to refer to personal experiences, and imply a possibly important association between the information seeking targets and real world attributes.

2.2 Displaying Searching Criteria

The presentation of the searching elements in the search interface has long been a research question. [5] took an insight into the attributes recalled, explored the characters of false and partial recall, and gave suggestions of elements to include in PIM interfaces. For example, as visual elements from other pages apart from the first page retrieved are equally remembered, the preview result should not be limited to the first page. Also, they noted that subjects tend to partially correctly recall file paths, dates, so they suggested giving users more flexibility for inputting such information, such as that used in [14] for the attribute of time.

A more human cognitive oriented interface for displaying time is in *Memory Landmarks* [15] where photos indicating landmark events from real life were displayed as anchors to help a user locate the time of the target. Theories in temporal memory suggest that people represent 'time' in their memory as a series of events, and by estimating the distance from the events where time is tagged using symbolic names. Their user study showed that this interface significantly improved searching efficiency and user satisfactory level.

Many studies on PIM have argued for the crucial role of the individual's episodic memory. While the episodic memory suggests that when we retrieve information from memory, we experience 'mental travel back to that scenario'[16], 'seeing' pictures which may indicate the location, weather, light status, etc. This implies that we may be able to find better ways of presenting searching options which make good use of what a user can recall which they may associate with an experienced artefact, an event or item.

3. PROPOSED INTERFACE

We aim to develop a searching interface for personal archives, providing the user with query options based on what they tend to remember.

Most current searching interfaces provide users with the same options to input in searching queries at both the initial search stage and the result re-finding stage; anticipating that users will get some improved ideas for query feature in ever next trial search. Many well featured interfaces provide the users with a long list of options at the initial searching interface, while in most cases, users only use the basic options, e.g. keywords.

The narrative interface [17] is a good way of exploring this question without limiting user's thoughts, and gives them plenty of freedom and ease to follow what they recall. However, users are not always dedicated, they will not be happy to type in every thing they can recall, and of course, the gap between natural language and information systems must still to be handled. Besides, sophisticated users may only input what may be good for

searching. To balance the information searching needs and the user effort, we want to equip the interface with simple enough initial searching input fields which allow free narrative text input, with advanced and more structured searching options in a re-finding panel in a screens presenting results of initial searches, where different but rich searching options will be provided. The nature of a re-finding interface is actually a search-aided browsing application. We plan to adopt the approach taken in the 'Memory Landmarks'[15] to display landmark items (both events in real life and those in the digital world) on a timeline to help users, and provide more search options from the items listed, e.g. allow searching from these items via certain associations. More importantly, advanced options should elicit search features from users that they tend to recall and make these available in the interface. The proposed evaluation strategy is discussed in Section 5.2.

4. WORK IN PROGRESS

4.1 Exploring possible searching options

As our previous pilot study with a single subject has little external validly regarding personal differences, we plan to carry out a diary study with more participants, in an even less intrusive way. We are particularly interested to know what currently omitted potential features could be exploited. A small scaled study has been carried out since February 2009, and will last for 2 months. We have tried not to restrict the subjects' thoughts to the options which current search systems provide. To achieve this we gave them two examples and instructed them not to limit themselves to what the examples suggest. In this study, the participants were basically required to take notes of the information and files that they tend to look for, and submit this using an online form thereafter when the Internet access is available. The data collection is anonymous.

4.1.1 Participants

Four participants volunteered for this study, with three outside our research group and described themselves as not being good at searching information online, and one being myself.

4.1.2 Results so far

11 cases were collected in the first week, indicating that in future that the participants will need to be encouraged to make more data contributions.

Due to the limitation of free recall, and the personal differences in deciding which activities are suitable entries for the diary, e.g. one participant only tended to record those that they can recall many details for or which are really difficult tasks. However, none of the cases above included those that they remember very well the keywords, or the associated filename or title, which could be used directly in current searching interfaces. In fact, the successfully recalling of precise details (e.g. file name, some keywords, and titles) is not a rare case in real searching tasks. However, 8 out of 11 of these cases have associated information recalled, either from the real world (e.g. conferences) or digital files and activities (e.g. emails, documents, information searching tasks). This is congruent with what Blanc-Brude found in [5], and that in our preliminary single case study [13]. It implies a potential interesting application of utilising associated information to assist user searching for items in personal life achieves.

4.2 Keywords users tend to use

One of our planned efforts on information presentation to explore the use of content-based keywords in searching, and in particular, the mistakes uses tend to make, so as to develop corresponding assistant features for the search by keywords options (fields such as file name, title, keywords in the document).

The result in [18] indicated to us the fact that a user's searching query may change overtime even for exactly the same task, and that the words they use may not be exactly those appearing in the target item that they are looking for. According the psychology theories on learning and memory, we assume that the use of keywords may change due to the user's knowledge (esp. vocabulary) evolving with experience, and that people tend to use words that they are currently familiar with, depending on the recency and frequency, etc. of encountering that word/phrase ([19] for more explanations). We want to explore the errors that users tend to make when generating content-based keyword queries, so as to develop strategies to assist their free text input accordingly. To explore this, we are doing a small scale pilot study in natural a setting over a 2 month period.

4.2.1 Participants

Four participants outside our research group volunteered for this study. They are all postgraduate student in computing or electronic engineering.

4.2.2 Material

To make it in a more natural setting where the subjects actually search for things, Google desktop search was used to index their data on their working computers, and an interface similar to Google desktop's pop out searching window was developed to let subjects search into their Google desktop database. The application recorded every instance of queries entered for a search task as well as the full text of the searching target and its filename/ page title are recorded.

4.2.3 Procedurals

Subjects were instructed in using this searching tool to search for any files or webpages they have visited. They were also told what information the application captures, so that they can decide when to use this search application, and could use a standard search application at other times. They are allowed to search several times until they find the target or decide to give up.

4.2.4 Results so far

The results so far (two weeks) are not as promising as expected. As most of these participants' time was spend on coding, their searching actions with this application are mostly searching for a code file using the filename. We plan to expand the study to research students from wider backgrounds, especially those who need to undertaken plenty of reading.

5. PROPOSED STUDIES

5.1 Diary study

To minimize the problem of diary entry bias mentioned above, we should have the user decide to add an entry before they know how much they can recall, and thereafter how difficult the searching tasks will be. Thus the burden of going into the add entry interface should be minimized. The Google desktop Ctrl+Ctrl short cut key is a good example. We want the subject to start adding an entry as soon as they decide to do some information seeking in their data

collection, without evaluating how well they can do with such an entry.

In case sometimes people may forget to mention some related information even they can recall of, a cued-recall approach should be used [5].

In this study, we particularly want to explore what associated information that can be recalled, and how likely it is to be recalled correctly. The data collection will also be anonymous, so that participants will be comfortable entering some essential information that involves some privacy. We expect to get a minimum of 20 diary entries from each subject over a period of 1 month. For the features the subjects recall, we will need to decide which ones to use based on the feasibility of using current techniques to capture and exploit recalled features, and their actual value in IR. For example, subjects may always remember the geo-location, which is the only place they access their computer, this information would have no utility for IR since it is always the same. However, if the remembered geo-location is for the photos taken at difference places, it may be a very good searching cue.

5.1.1 Participants

We aim to recruit 20-40 participants from a wide range of backgrounds within the campus. Subjects will be from outside our research group and have little knowledge of our previous research, so that their later performances will not be biased by their knowledge of our previous work on the use of contextual data in search. Their experience and skills will be evaluated before the start, and we expect half the participants to be those who have less experience and knowledge of IR and desktop searching systems in general.

5.1.2 Material

We are currently in the process of designing and developing the diary interface, while taking into account outcomes of previous studies including [5] for the options of searching fields to include, We also need to give users with flexibility and freedom to add those types of recalled data which are not included in current interfaces. It will be a desktop application with a floating icon on top of the screen, so that (I assume) this icon can remind the users to add entries to the diary when an information-seeking task comes to them.

The layout of the diary searching-link interface will give the interface panels generally equal opportunity to be clicked by the user randomly, so that it won't bias the subjects' recall performance of the above information, e.g. presenting the panel of physical context at the very best eye-catching position, subjects may tend to enter such information more carefully. The last step before completing a diary entry will seek to elicit some additional information which may be contribute to potential factors which influence the recall performance, including:

- Details of overall memory of the target
- Frequency of accessing the target item
- Time elapsed from last time accessing the item

I anticipate that this experiment will provide some solid input for future explorations, including: the metadata that should be used to annotate the PL items, and in particular, the types of association people tend to recall, and the factors which influence recall of information.

5.2 System Evaluation

I plan to do system evaluation in real information seeking situations. That is, in a considerably longer period, e.g. 1 month, to encourage participants to use it when they want to look for something. Due to the uniqueness of the data collections, e.g. in relation to one's own memory, and privacy issues, this will require the participants searching into their own data collection.

5.2.1 Participants

Due to the uniqueness of data collection regarding its connection with individual' own memory and other issues privacy concerns, the participants must be tested on their own PL data collection. Thus, there is a big challenge to get enough participants with data collections since they must be gathered over an extended period, and this requires considerable commitment from each participant and availability of suitable hardware. We currently have three life loggers actively engaged in collection of personal data over a one-year data collection period.

5.2.2 Data collection

Based on the types of data needed in the system described in the next section, we need data collections including: logs of computer activity, rich digital photo collection, and physical contextual information such as geo-location. The current data collection from the three life loggers includes:

- All their computer activities as well as logs from their mobile phones: Recording using applications such as Microsoft Digital Memories [1] and the Slife [2], we are logging every foreground window activity and storing full text information of items (if available).
- Microsoft SenseCam images which has been passively taken around every 20 seconds during daily life
- Geo-location data from GPS devices.
- Bluetooth records: recording people and objects.
- We also expect to include other data types which the diary study results may indicate.

Due to occasional equipment failures and data availability all of this data streams have some small gaps.

5.2.3 Proposed Methods

A long term implicit observation of natural searching tasks with this interface will need to be tested. Although the results can be achieved by comparing the performance on this interface with a baseline one, it is almost impossible to do between subject studies by assigning 3 participants to 2 groups, nor is it feasible to do within subject study by forcing subjects to switch between interfaces.

6. CONCLUSION

Current searching interfaces for personal lifelogs usually provide searching options based on what the current system provides. This may omit some potentially feasible options users tend to

[1] http://research.microsoft.com/en-us/collaboration/focus/cs/memex.aspx

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

remember well. In this study, I aim to explore the omitted types of information, with special focus on exploiting the association between items in a PL, and exploring the use of these items in a way that people represents it in their memory.

7. ACKNOWLEDGMENTS

This work is funded by a grant under the Science Foundation Ireland Research Frontiers Programme 2006, and supported by Microsoft Research under grant 2007-056.

8. REFERENCES

[1] Jones, G., Gurrin, C., Kelly, L., Byrne, D. and Chen, Y. Information Access Tasks and Evaluation from Personal Lifelogs. In *Proceedings of the EVIA 2008 - 2nd International Workshop on Evaluating Information Access, A Satellite Workshop of NTCIR-7* (Tokyo, Japan, 2008), Tokyo, Japan, 2008.

[2] SparckJones, K. *Comments on Grand Challenge Document (GCD): 'Memories for Life': managing information over a human lifetime*. Cambridge, 2004.

[3] Gemmell, J., Bell, G. and Lueder, R. MyLifeBits: a personal database for everything. *Commun. ACM*, 49, 1 2006), 88-95.

[4] Lee, H., Smeaton, A. F., O'Connor, N. E., Jones, G., Blighe, M., Byrne, D., Doherty, A. and Gurrin, C. Constructing a SenseCam Visual Diary as a Media Process. *Multimedia Systems*, 14, 6 2008).

[5] Blanc-Brude, T. and Dominique, L. S. What do people recall about their documents?: implications for desktop search tools. In *Proceedings of the Proceedings of the 12th international conference on Intelligent user interfaces* (Honolulu, Hawaii, USA, 2007). ACM, Honolulu, Hawaii, USA, 2007.

[6] Jaimes, A., Omura, K., Nagamine, T. and Hirata, K. Memory Cues for Meeting Video Retrieval. In *Proceedings of the Proceedings of the the 1st ACM workshop on Continuous archival and retrieval of personal experiences* (New York, New York, USA 2004), New York, New York, USA 2004.

[7] Mor, N., Susumu, H., QianYing, W., Hector, G.-M. and Andreas, P. Context data in geo-referenced digital photo collections. In *Proceedings of the Proceedings of the 12th annual ACM international conference on Multimedia* (New York, NY, USA, 2004). ACM, New York, NY, USA, 2004.

[8] Järvelin, K. and Wilson, T. D. On conceptual models for information seeking and retrieval research. *Information Research*, 9, 1 2003), paper163.

[9] Robert, G. C. I. and Manuel, P.-Q. Factors and Evaluation of Refinding Behaviors. In *Proceedings of the SIGIR 2006 Workshop on Personal Information Management* (Seattle, Washington, August 10-11, 2006, 2006), Seattle, Washington, 2006.

[10] Elsweiler, D. and Ruthven, I. Towards task-based personal information management evaluations. In *Proceedings of the Proceedings of the 30th annual international ACM SIGIR conference on Research and development in information retrieval* (Amsterdam, The Netherlands, 2007). ACM, Amsterdam, The Netherlands, 2007.

[11] Michael, H., Holger, H. and Ron, R. Cognitive Principles for Information Management: The Principles of Mnemonic Associative Knowledge (P-MAK). *Minds Mach.*, 17, 4 2007), 445-485.

[12] Kim, K.-J., Jung, M.-C. and Cho, S.-B. *Episodic Memory for Ubiquitous Multimedia Contents Management System*. Springer Berlin / Heidelberg, 2007.

[13] Fuller, M., Kelly, L. and Jones, G. J. F. Applying Contextual Memory Cues for Retrieval from Personal Information Archives. In *Proceedings of the Proceedings of Personal Information Management, Workshop at CHI 2008* (Florence, Italy, 5-6 April 2008, 2008), Florence, Italy, 2008.

[14] O'Hare, N., Gurrin, C., Jones, G. J. F., Lee, H., Noel, E. O. C. and Alan, F. S. Using text search for personal photo collections with the MediAssist system. In *Proceedings of the Proceedings of the 2007 ACM symposium on Applied computing* (Seoul, Korea, 2007). ACM, Seoul, Korea, 2007.

[15] Cutrell, E. Search User Interfaces for PIM. In *Proceedings of the Personal Information Management - A SIGIR 2006 Workshop* (2006), 2006.

[16] Tulving, E. Episodic memory and common sense: how far apart? *Philos Trans R Soc Lond B Biol Sci*, 356, 1413 (Sep 29 2001), 1505-1515.

[17] Gonçalves, D. and Jorge, J. A. In search of personal information: narrative-based interfaces. In *Proceedings of the Proceedings of the 13th international conference on Intelligent user interfaces* (Gran Canaria, Spain, 2008). ACM, Gran Canaria, Spain, 2008.

[18] Kelly, L., Chen, Y., Fuller, M. and Jones, G. J. F. A study of remembered context for information access from personal digital archives. In *Proceedings of the Proceedings of the second international symposium on Information interaction in context* (London, United Kingdom, 2008). ACM, London, United Kingdom, 2008.

[19] Chen, Y. and Jones, G. Integrating Memory Context into Personal Information Re-finding. In *Proceedings of the FDIA 2008 - 2nd Symposium on Future Directions in Information Access* (London, UK, 2008), London, UK, 2008.