Memory Support for Desktop Search

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

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

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

ABSTRACT
The user's memory plays a very important role in desktop search. A search query with insufficiently or inaccurately recalled information may make the search dramatically less effective. In this paper, we discuss three approaches to support user's memory during desktop search. These include extended types of well remembered search options, the use of past search queries and results, and search from similar items. We will also introduce our search system which incorporates these features.

Keywords
Desktop search, memory, suggestive interface, diary study.

1. INTRODUCTION
Desktop search refers to information seeking in personal archives, which include one's emails, documents, visited web pages, digital photos, mp3 file, and mobile phone text messages. The variety and amount of items in personal archives continues to increase with the development of new computing and storage technologies. The increased complexity and size of personal archives means that more advanced desktop search techniques are needed. Since personal archive items are usually downloaded, received, created, edited or viewed (read) by the individual owning the personal archive, desktop search targets are usually what one has encountered previously. Therefore, they have some links with one's memories associated with the items. When one looks for things in one's personal archive, the approach and queries one uses may depend not only on current task context, but also rely on what one can remember about the target you are seeking. For example, with windows desktop search [1], one usually needs to recall at least one type of information about the target(s), such as the filename, or the last visiting date.

There are usually two stages for desktop search: the first is to determine what target to look for, then the second is to look for (search for) this target. For example, when I want to look for the time of a meeting later this week, I first need to know where I can find it, or where I saw such information before. After I recall that I encountered this information in an email, I will then need to recall information such as when I received this email, who sent it to me, or the subject of the email, in order to find it out from my mail inbox. If I can't remember any of above information it may be very difficult for me to find this email and the information. This example indicates the important role of the data owner's and user's memory in performing re-finding tasks.

In this paper, we describe three memory feature driven approaches to assist desktop search (in sections 2, 3 and 4 respectively). In section 5 we present our desktop search system which embeds functions corresponding to these three approaches. Finally we give a brief overview of an experiment we are undertaking to test these approaches with our desktop search system.

2. USE WELL REMEMBERED INFORMATION TO SEARCH

2.1 Related Works
While it has been found that people usually prefer using simple queries and series of small steps to narrow down the dataset and approach a search target [1], entering queries to retrieve the results directly is still an important approach. There have been several studies looking at utilizing people's memory features in search (e.g. [2], [3], [4]). Most of these studies believe that the key to using memory features to support desktop search is to know which features of the items people tend to remember. Enabling users to search with likely remembered features of items is of course an important way to improve desktop search efficiency. However, this is not all. In fact there is usually another step, browsing the retrieved results to locate the target or the precise piece of information that is needed. Works looking at this aspect include [3, 5, 6], etc. In the study by Ringel et al. [6], they enabled users to browse the result on a temporal dimension together with items representing personal and public important events. They found that search times were reduced significantly when the user had access to episodic context. This implies that people's memories about their visited items (items in their personal archives) are not isolated units which are comprised only of the memory of the attributes of specific items, but rather that they are associated with the episodic context of accessing these items. In preliminary studies we also found that a subject had more reliable memory of episodic context (e.g. location) than of the target items themselves. Search queries which combined content and context information showed greater advantage over long term [7]. However, the result of this experiment was only from one subject. For this reason, we are conducting a diary study to explore what other people remember about their personal archive items when they look for them.

[1] http://www.microsoft.com/windows/products/winfamily/desktopsearch/

Copyright is held by the author / owner(s)

SIGIR'10, Workshop on Desktop Search, July 23, 2010, Geneva, Switzerland.

2.2 Diary Study

2.2.1 Participants

This is an on-going study. We have so far completed a pilot stage with four subjects. Many more participants have agreed to participate in this experiment. All the participants were university research students majoring in computer science. They were invited to participate in person. The details of the diary study were explained to them before they signed up.

2.2.2 Material

Diary books with printed questionnaires (shown in figure 1) were given to each subject. Each of the diary books contains 40 pages each of which can hold 16 diary entries. The other 8 pages include a participating consent form, instructions, and two blank pages for comments. On one side of each paper, there is a field for description of the target, one big field for free recalled details related to target, and multiple choice questions about the type of the target if the target field is not filled in because it contains private information, the frequency of access, and remembered occasions of access. On the back of each page there are multiple choices questions asking about their reasons and approaches of this re-finding activity, and a group list of episodic context features from which to select their remembered ones, such as "your location", "people around you", "other digital items accessed around that time". This is because people may forget to list some information during free recall even though they remember the details of them. An electronic version of this diary book is also available online.

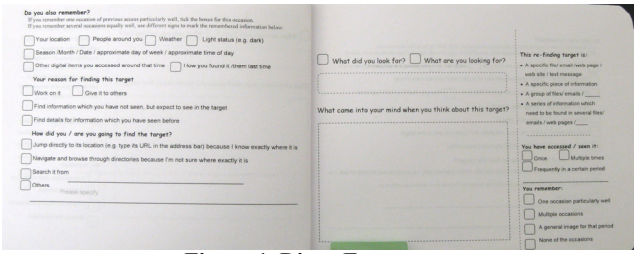

Figure 1. Diary Entry pages

2.2.3 Methods

The participants were asked to fill in a diary entry when they are looking for or after they found digital items which they have encountered before. One week after they were given the diary book, each of them was visited to interview them about the diary entries they have written by that time. This was to check if they fully understand the questions, inspire them to recall more aspects of the target related information, and encourage them to add more high standard diary entries more frequently.

2.2.4 Results so far

To date, four participants have completed the study and returned the diary book. Due to the small number of participants who have so far completed the study, statistical conclusions about remembered features cannot be provided for this paper. However, the following is a summary of our findings at this stage. For 91% of the diary entries, the participants claim to remember at least one occasion of interacting with the targets, though for 60% of these they only remember a general context. For example, one recalled "I was working on it day and night to beat the deadline". People sometimes also remember why they accessed that item previously, associated events or tasks, or people involved in those events. This suggests that if their remembered episodic information of interacting with the item can be recorded, it can be used in subsequent search. Another interesting finding is that they claim to remember how they found the target previously, sometimes even remembering the exact queries used to find the item. This is consistent with existing studies which show that we sometimes search for things which we have searched for before, and use similar queries [8].

2.2.5 Discussions

We believe that enabling users to search with likely remembered features of their targets and episodic features of previous access of these targets can make desktop search easier for the user. Admittedly, not all types of well remembered information can be recorded and translated to textual content to perform search. For example, what one thinks, smells cannot generally be captured. However, searching by these types of information is not impossible if past queries and results can be used given they searched by these types of information before. In the next section, we discuss a search approach using previous search queries and retrieval results.

As for the question of which episodic context information should be used, we will further explore this in our diary study and a possible follow up survey to statistically investigate the most likely remembered features.

3. USE OF PAST QUERIES AND RESULTS

3.1 Background

According to the findings of our diary study, people sometimes remember how they found information or items last time. We propose that an additional background search into a data collection of one's past desktop search queries using the current search query can assist the user to more efficiently retrieve a currently sought item, if they remember part of a previous query and are looking for the same item this time. This is because individuals may use their remembered previous queries to search again. Although the same query may bring them exactly same result list, they may still need to browse for the exact result in the retrieved list. Therefore, each record in this data set should include not only the past query, but also the result which was found to be relevant with this query. When a matching query is found, the corresponding result (item) can be included in the result list for the current query. If this happens to be the item that the user wants, it can save them the effort to browse for it in the list. The Heystaks tool [9] can also serve this function, however it focuses on sharing past queries in a social community, which is different from personal desktop search. Besides, it is limited to only one search field. Desktop search usually has many more potential search options. .Moreover, , people may not always remember correctly which queries brought them the target last time if they tried several queries in rapid succession. This is because that when we recall something, we actually reconstruct the whole thing or story from atomic pieces of associated information [10]. Omitting or misplacing any pieces of memory may lead to mismatching queries. Current major desktop search tools filter the data collection and return results which match *all* search criteria. Thus, a memory mistake of a single feature can cause the search to fail. We believe that if the search system can provide relevant items which match

"any" of the multiple search criteria, such memory failure causing information retrieval problem can be reduced. This is because even if the user recalls a piece of information incorrectly for one field, the potential target may still be retrieved because it matches other criteria.

3.2 Solution

For a similar reason, people may remember how they found certain items previously, and may even remember the queries they used. However, if someone developed a query iteratively to locate the desired target, they may not necessarily remember correctly which query actually retrieved the target for them, since they generated these queries in almost the same context and the queries they used may have equal-strength links to the experience of achieving the target, if the person did not intend to learn which one is the "right" query.

Therefore, we believe that to fully utilize memory of previous search experiences, and supplement possible memory faults, not only the query which brings a "click" on the results should be recorded into the data collection of past queries, but also other queries for that same search target should be recorded together with the selected target items.

Also, if the main data collection (index of personal archives) has multiple fields and the search system supports queries of multiple fields, the index for past queries should be indexed to the corresponding fields. Similar to the situation discussed earlier when retrieving documents associated with past queries and results, items with any matched field should be considered in case the user only partially remembers the previous query.

4. SUGGEST FOGOTTEN TARGETS

The above methods focus on the second step of desktop search, that is, recalling information related to the target to search for the target. However, sometimes people do not even remember the existence of possible targets, and therefore the generated queries may not match any features of these potential targets to retrieve them. Or sometimes the image of the potential target item is too blurred and the user is unable to describe what this item is with any of the provided search options.

We believe that searching by similar items can assist people in finding these potential targets which are less well remembered or even forgotten. For the former type of potential targets which are omitted because the user forgot about its existence at the time of searching, a re-search by features from an already-found target can possibly bring up these items.

As for the later type, since it is usually much easier for familiarity-based recognition than recalling exact details, it is possible that its features can be recognized as those of a potential target if they are presented, even if they are displayed as features of some other items [11]. In other words, when the searcher sees an item which is similar to the target in certain respects, these features may be recognized as a target even if the searcher is not sure which exact features can be used to generate queries to search for the target.

4.1 Challenges brought by these approaches

While the above approaches aim to increase the chance of retrieving the correct result and improve recall, they may equally bring more noise, and reduce the precision of the search. Thus they bring the challenge to information retrieval techniques of pushing potentially relevant targets to the top of the retrieved list. Refined search may also be needed after an initial search. The difficulty is how to support users in performing a refined search with less noise and possibly better recall, and more importantly to improve the precision. One common solution is to use filters when browsing results. However, filtering results by certain criteria may not significantly reduce the available result set. This is because people usually only select information they are certain about (possibly well remembered) as filters, so they may have used these criteria as a major part of their query too. Thus it is very likely that most of the results may satisfy these filter criteria. More efforts are needed to design browsing functions to facilitate locating information in large result sets.

5. OUR SEARCH SYSTEM

In this section we will introduce our search system which embeds the memory support functions outlined above. This system is primarily designed for accessing extended personal archive data collections. Such datasets not only include digital items that one has encountered, but also timestamps for accessing them, and the physical world context when accessing them, such as the person's location, or the weather. With this system, we can test whether enabling search by these types of episodic context information can improve the effectiveness of desktop search for users.

5.1 IR algorithm- multiple fields

The underlying search algorithm enables search of multiple fields with "or" logic as described in section 3.1. An in-house developed version of BM25 [12] for Lucene[2] which uses fielded search is used in our experiments. This approach sums individual field scores to arrive at the overall retrieval score for an item. For each individual item field a 'should occur' query clause is used on query terms [13]. The retrieval approach used means that only one result set will be returned for each search action. On the search interface, tags are used to enable users quickly jump to wanted search fields and remind them of possibly remembered options.

5.2 Index of past queries

To make use of our "remembered" past queries (as proposed in section 3), the system logs every search query and results selected from the search interface. A click on the search button is defined as a single search action. A search session includes all search actions for the same target(s). Queries for each search action are recorded and indexed into a Lucene index as a single document. Each document contains a field for task_id, fields for each search field, including fields for the title, keywords, item type, etc., and a field for the item_ids of targets found during the entire search session.

When a user searches, the query not only searches in the Lucene index of the desktop data collection, but also searches in the index of past queries The results retrieved from both indexes will be presented in an integrated fashion, so that items appearing in both lists are merged to reduce the total amount of information being presented to the user. Items retrieved from the past query index will also be marked so as to give the user an

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

indication that this item is likely to be what the user wants selected previously when entering this query. The question of how to rank the results retrieved from the past query index is still to be explored.

5.3 Search by Similar items

This system is also designed to allow search from result items, hoping to bring more potential qualified items according to the rationale we explained in section 4. Users can click the result to look for items which are similar in content (file content, author of the email or SMS, type, etc.) or adjacent in time (as shown in Figure 2). We are also planning to enable search by physical

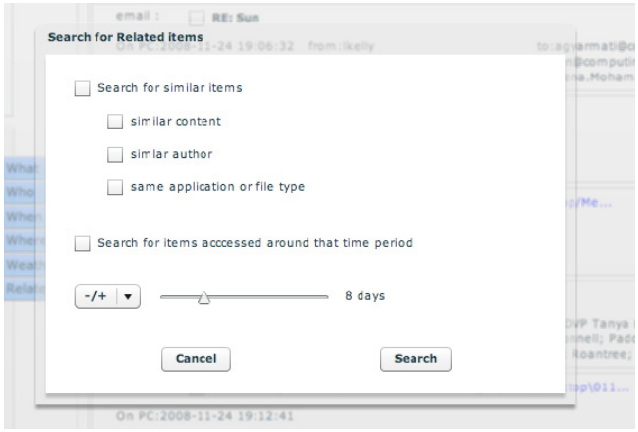

Figure 2. Screen cut of our desktop search interface: search from similar items

context of selected result items, e.g. similar weather, place, surrounding people. However, the elements which make the physical context look similar may not always be easily describable. For example, visual presentations (e.g. a photo) may be much more powerful in triggering people's memory of the physical context of the target item.

6. CONCLUSION

In this paper, we proposed three typed of approaches to support the user's memory imperfectness during desktop search. Firstly, based on a review of related work, and the findings in our on-going diary study, we suggest that episodic context can be recorded and exploited in search, and past queries can be utilized because people may remember previous queries to search again. The second approach focuses on supplementing memory by retrieving from an index of past queries to retrieve previously selected targets for the entire search session. This is because that people may not necessarily remember exactly which query brought the target to them previously. The third approach aims at overcoming memory problems at the first step of re-finding, that is, identifying the potential target. We hypothesized that search from similar items can reduce the search problems caused by memory failures at this step. We also introduced our desktop search system features which support these approaches. While all these approaches potentially enable more effective search for relevant targets, they may equally bring more noise. We are currently conducting an experiment to explore whether these approaches can improve desktop search performance, e.g. retrieve more relevant item, or reduce the time and effort spend on each search task. We expect to be able to report the results at the workshop.

7. ACKNOWLEDGMENTS

This work is supported by grant from Science Foundation Ireland Research Frontiers Programme 2006. Grant No: 06/RFP/CMS023.

8. REFERENCES

[1] Teevan, J., *et al.*, "The perfect search engine is not enough: a study of orienteering behavior in directed search," In the Proceedings of the SIGCHI conference on Human factors in computing systems, Vienna, Austria, 2004.

[2] Dumais, S., *et al.*, "Stuff I've seen: a system for personal information retrieval and re-use," In Proceedings of the 26th annual international ACM SIGIR conference on Research and development in informaion retrieval, Toronto, Canada, 2003.

[3] Elsweiler, D., *et al.*, "Exploring memory in email refinding," *ACM Trans. Inf. Syst.,* vol. 26, pp. 1-36, 2008.

[4] Blanc-Brude, T. and Scapin, D. L., "What do people recall about their documents?: implications for desktop search tools," In Proceedings of the 12th international conference on Intelligent user interfaces, Honolulu, Hawaii, USA, 2007.

[5] Freeman, E. and Gelernter, D., "Lifestreams: a storage model for personal data," *SIGMOD Rec.,* vol. 25, pp. 80-86, 1996.

[6] Ringel, M., *et al.*, "Milestones in time: The value of landmarks in retrieving information from personal stores," in *Interact 2003, the Ninth IFIP TC13 International Conference on HCI*, Zurich, Switzerland, 2003, pp. 184--191.

[7] Kelly, L., *et al.*, "A study of remembered context for information access from personal digital archives," In Proceedings of the iIIX, London, United Kingdom, 2008.

[8] Teevan, J., *et al.*, "History repeats itself: repeat queries in Yahoo's logs," In Proceedings of the 29th annual international ACM SIGIR conference on Research and development in information retrieval, Seattle, Washington, USA, 2006.

[9] McNally, K., *et al.*, "Towards a reputation-based model of social web search," In Proceeding of the 14th international conference on Intelligent user interfaces, Hong Kong, China.

[10] Loftus, E., "Memory Distortion and False Memory Creation," vol. 24, ed, 1996, pp. 281-295.

[11] Cleary, A. M., "Recognition Memory, Familiarity, and Déjà vu Experiences," *Current Directions in Psychological Science,* vol. 17, pp. 353-357, 2008.

[12] Robertson, S., *et al.*, "Simple BM25 extension to multiple weighted fields," In Proceedings of the thirteenth ACM international conference on Information and knowledge management, Washington, D.C., USA, 2004.

[13] Kim, J., *et al.*, "A Probabilistic Retrieval Model for Semistructured Data," in Proceedings of the 31th European Conference on IR Research on Advances in Information Retrieval, Toulouse, France, 2009.