Information Access Tasks and Evaluation for Personal Lifelogs

Gareth J. F. Jones Cathal Gurrin Liadh Kelly Daragh Byrne Yi Chen
Centre for Digital Video Processing
Dublin City University, Glasnevin, Dublin 9, Ireland
gjones@computing.dcu.ie

Abstract

Emerging personal lifelog (PL) collections contain permanent digital records of information associated with individuals' daily lives. This can include materials such as emails received and sent, web content and other documents with which they have interacted, photographs, videos and music experienced passively or created, logs of phone calls and text messages, and also personal and contextual data such as location (e.g. via GPS sensors), persons and objects present (e.g. via Bluetooth) and physiological state (e.g. via biometric sensors). PLs can be collected by individuals over very extended periods, potentially running to many years. Such archives have many potential applications including helping individuals recover partial forgotten information, sharing experiences with friends or family, telling the story of one's life, clinical applications for the memory impaired, and fundamental psychological investigations of memory. The Centre for Digital Video Processing (CDVP) at Dublin City University is currently engaged in the collection and exploration of applications of large PLs. We are collecting rich archives of daily life including textual and visual materials, and contextual context data. An important part of this work is to consider how the effectiveness of our ideas can be measured in terms of metrics and experimental design. While these studies have considerable similarity with traditional evaluation activities in areas such as information retrieval and summarization, the characteristics of PLs mean that new challenges and questions emerge. We are currently exploring the issues through a series of pilot studies and questionnaires. Our initial results indicate that there are many research questions to be explored and that the relationships between personal memory, context and content for these tasks is complex and fascinating.

Keywords: *lifelogging, human digital memories (HDMs), personal information management, user information needs, personal lifelog access evaluation*

1 Introduction

Early information retrieval (IR) evaluations were concerned with library records and academic sources, while much work on summarization (or abstracting) focused on scientific documents. More recently the development of standardised evaluation IR workshops such as TREC, CLEF and NTCIR focused initially on retrieval from newspaper document collections. Subsequently they have expanded the scope of their investigations to other tasks which have individual characteristics. For example, web data, patent and legal collections, and multimedia content. Work on information extraction and summarization, such as that reported at MUC and DUC workshops has similarly focussed on well defined published texts such as news reports. A common characteristic of all of these data sources is that, subject to suitable distributional agreements with copyright owners, this data can be distributed to the signed up participants of the evaluation exercises. Further it has generally been possible to capture and study the information needs or search requirements of expected users of these types of collections. These have then been used to develop experimental search topics or other user requirements to enable exploration of the task effectiveness for these collections. The success of these systems then being assessed based on manual post hoc assessment of the data by analysing documents retrieved or submitted for assessment by workshop participants. This framework for evaluations makes a number of assumptions about the way in which the data can be used and the test evaluation collections designed.

While these methods work well, albeit with limitations, for these information sources which are in some sense published, formally structured or publically available, other types of collections introduce new challenges for the design evaluations for information access tasks. Of particular interest to this paper is the emergences of Personal Lifelogs (PLs), sometimes referred to as Human Digital Memories (HDMs)[1].

[1] Although referring to them as HDMs is somewhat misleading

Recent developments in data capture and storage technologies mean that it is now becoming possible to capture and store electronic records of many of an individual's life activities. PLs can most obviously store materials such as emails that an individual has sent and received; text documents from the web and elsewhere that they have read, written, or downloaded; multimodal footage from their life experiences such as photographs taken or received, videos recorded or viewed, and music created or listened to. But can also include details of places visited derived from location data (e.g. harvested from GPS sensor tracklogs), details of people met and objects present (e.g. via Bluetooth sensors), and details of the subject's physiological state (e.g. via biometric sensors) when information was encountered. This information can be captured from a range of increasingly ubquitous devices such as personal computers, mobile phones, cameras, video recorders, audio recorders, personal GPS sensors, Bluetooth sensors embedded in various devices, and small portable biometric sensors.

A number of research projects are underway exploring the collection of large heterogeneous PLs, probably the best known of which is the Microsoft MyLifeBits initiative [5]. While the development of integrated PLs is still not commonplace, many individuals are already collecting personal archives of one or more of these data streams from their lives. It is quite reasonable to expect such archives to grow massively in size, scope and complexity in the coming years. The ability to gather and store all this personal data is a considerable achievement in itself. However, it is probably only worthwhile if these collections can be usefully employed for some purpose. Fortunately such archives have many potential applications including search activities helping individuals recover partial forgotten information, sharing experiences with friends or family, telling the story of one's life, clinical applications for the memory impaired, and fundamental psychological investigations of memory. In order to provide these applications in an effective way new approaches to information access and extension of existing ones are required, and in order to understand their behaviour and effectiveness, issues of evaluation need to be addressed and applicable test procedures defined. To date little work has been appearing which addresses this need for new evaluation strategies.

In this paper Section 2 first examines the similarities and differences between PL and more traditional informaiton collections, Section 3 looks at some of the tasks for which PLs might be applied, while Section 4 describes some of the metrics which can be used to measure performance over these tasks, Section 5 summarises work in PL applications to date, Section 6 describes current work in our group at Dublin City University, and finally Section 7 concludes with some thoughts about future initiatives in evaluation activities for PLs.

in terms of the accepted definitions of memory.

2 Personal Lifelogs

Information access applications for PLs present new challenges and opportunities for technologies and evaluation. In this section we explore the differences between PLs and more traditional content archives. A key feature of PLs is that they will often incorporate a combination of many types of media, including audio, video, images, and many types of textual content. There is also the potential for a large percentage of noisy data in these archives. This can arise since much of the content is captured passively by computing devices without active participation by the user. The result of this approach to capture will mean that while there will be high coverage of potentially interesting and important life experiences, it is also likely that much of the captured material will be of no future interest to anyone. In addition, since people often return to the same information many times, and in the case of some items they will be revised incrementally over time, many items in the archive may be very similar, repeatedly covering the same topic. Other differences between PLs and traditional collections include that the user may have forgotten about an item or event, and they may not even be aware that a particular piece of data was captured and is available in the PL.

When performing information access tasks from within a PL users may not be able to describe clearly what they are looking for. In many cases items may not have formal textual descriptions, meaning that they cannot be retrieved using standard text or meta-tag based retrieval methods. Another is that items may not be joined by inter-item links, meaning link structure cannot be utilized in the retrieval process. However, one feature with intriguing implications for information access is that items can be linked based on a variety of content and context attributes. All these features mean that there are new technical challenges associated with the PL domain.

One further crucial difference is that a PL is obviously *personal* to the individual collecting the data which raises many issues. Since they record the personal experiences of the individual whose life they capture, this user will often have unique memories associated with the items in the archive. While it is acknowledged that memory and various cues play a vital role in retrieving an item from human memory, current IR systems make little attempt to exploit what people remember about items. For example while an individual may work through the following thought "I remember finding the file on my laptop before when I was in the cafe drinking coffee with Pete when it was raining about a month ago" and use some of this information to find the file, we are not aware of any current

IR systems which allow a user to search for an item based on a complex combination of features including such details as the people in their vicinity or the prevailing weather conditions.

The rich variety of context data sources that are potentially available in PLs makes a compelling case for the exploration of their exploitation to improving the accuracy in information access applications. Beyond this observation, coupled with rich context sources is the additional information that can be obtained about items by linking them to related content. In standard IR, the exploitation of the web's linked structure has proven to be a important component in high quality web page retrieval [30]. The link structure is used to determine the quality and potential relevance of a page by looking at the number and quality of pages linking to it. It is interesting to consider whether exploitation of existing or inferred links within a PL could also be used to enhance the performance of information access applications. These links may be either content (e.g. linking based on the content similarity of items) or context (e.g. linking temporally based on user access patterns) based. To date only very limited work has been reported within PLs, for example [36] where result sets were linked based on user access patterns and the resulting linked structure was used to rerank the result set using PageRank type algorithms. We believe that there is significant scope to extend these ideas to new methods for information access from PLs.

Having considered some of the important distinctive features of PLs we next turn our attention to considering what user tasks they might actually be used to support.

3 Information Access Applications for PLs

We envisage PLs being used as part of a range of information access applications, some really being straightforward extensions of standard IR to a new information source, others being modifications or enhancements of existing user activities, and others being potentially new tasks. Some specific possibilities making use of a number of underlying information access technologies include: standard interactive search; proactively reminding the user of related experience from the past (e.g a previous visit to the same place), an advanced application of this sort might take the form of a type of prosthetic extended memory; allowing users to reminisce about experiences (a sort of personal browse or surfing service); social use (where the user might share their experiences with colleagues, friends and relatives) or, in a rather more complex scenario, multiple users might link sections of their PLs to create "community" memories; and their use in therapeutic treatment of memory impaired patients. Another strand of applications, which we don't explore further here, is their potential application to basic psychological memory studies, where they could be used to explore areas such as: success and failure of recall from memory, errors in memory recall, false memory syndrome, and learning and rehearsal studies.

Realising some of these possibilities is relatively simple, and indeed work to develop several of them is well underway, while others are much more ambitious, requiring extended periods of research. However, if they are to be explored in a scientific manner, they need to be scrutinized carefully in an attempt to properly understand the basic scientific questions and issues raised, and to develop appropriate experimental strategies to evaluate them. In order to do this we need to consider the available sources of data, classify it into well specified types depending on its features, and consider the means of capture. In terms of experimental design we need to be aware of the issues associated with developing evaluation strategies for information access tasks with traditional information sources, but also be aware of the unique features of PLs. These include issues such as the fact that individuals will often consider their PL data private to themselves. Thus the data cannot be distributed to researchers (as is generally the case for other experimental information access search data sets), and since much of the meaning and interpretation of content will also be personal, the individual will need to be involved in developing the test data for their PL and interpreting the results. Even where the user can at least to some extent be separated from the these activities, many people will not be willing to allow users to look at data from their PL even for the purposes of tasks such as relevance assessment. The requirement that the "owner" of the PL must be heavily involved in the collection of the data, the development of test data and the evaluation of experimental results, will necessarily limit the number and types of people who are able and willing to be involved in these activities. It is perhaps likely that people willing to make this level of commitment will not be properly representative of typical user groups, meaning that one has to design and interpret experiments with care. Also the extensive involvement of the individual may affect their behaviour at any stage of the experimental process, which may again call into question the validity of experimental results. For example, their behaviour during data collection may not be typical, and repeatedly gathering a range of experimental results may affect their knowledge of the collection restricting their ability to behave in a normal way when interacting with the data or biasing their interpretation of results.

In addition to the personal use of PLs, alternative scenarios are clearly possible, for example one user could give another permission to access their PL, perhaps within a family or among relatives making their PL available to their descendents. While the possibil-

ity to explore the life history of a grandparent directly from their PL offers fascinating prospects for the future, assuming grandparents are willing to be quite so open about their experiences, at present such historical collections do not exist, and our experience is that users are not willing to give this level of access to their associates. At this point, third party use of PLs has largely been outside the scope of the experimental studies.

At this point we turn our attention to consideration of some specific tasks in a little more detail.

3.1 Ad Hoc Search

One class of PL application is ad hoc search. Various scenarios are possible here. Users may perform a known-item search for a specific item or event which they partially remember, either to inform themselves of some forgotten detail or to share it with others. Alternatively, they may have a less well focussed information need where they either feel, know or believe that they have experience of something in their past, but are unsure of the details. Or perhaps, the user may wish to browse their PL with respect to some facet, such as experiences where person X was present. A more advanced version of this would have the system link together related materials, potentially dynamically forming further links in a query dependent manner, to provide users with suggested related memories. Some of these the user may remembered, but the significance of others may have been overlooked or they may have been forgotten entirely.

In our work we hypothesise that contextual information related with people's memory of the generation of and previous access to items and events stored in their PL can be used to improve subsequent access performance. Traditionally IR systems allow users to search based on their memories of keywords contained within items. However, over time individuals memories of these keywords can fade resulting in increased difficulty in locating the required information. In the case of some data such as files, it is often easier to remember non-textual elements such as location [3]. Non-textual memories of items may range from people present when accessing an item, to a phone conversation that interrupted your email, writing, etc. In our work on IR for PLs we are working to develop and test methods which integrate automatically recorded and derived context data types into traditional IR algorithms for improved retrieval in the PL domain [18]. Our results so far indicate that context can indeed be better recalled than content as the time from PL collection increases [27]. As part of this work we are seeking to better understand the processes and circumstances in which content and context are recalled accurately, inaccurately or forgotten, and to the extent to which these details can be generalised across users [12].

3.2 Narratives

In addition to traditional search activities, PLs offer new opportunities for other domains. One of particular interest to us is that of digital narratives. In their simplest form a user may wish to re-experience materials associated with some activity from the past. A slightly more formal application is the use of personal visual diaries captured using a passive camera, such as the Microsoft SenseCam [24], which regularly record images from the wearer's day enabling them to subsequently replay them. Visual diaries are a unique aspect of PLs that capture the actions of an individual from their perspective, thereby creating a personal archive containing a visual stream of everyday life. As an alternative to writing a diary, the visual diary or lifelog can be recorded from the start to the end of the day. Using a conventional diary, a person can easily review a day or any period of time quickly and efficiently, however with visual lifelogs, this process is more challenging because although the visual lifelog is a richer medium than a conventional diary, organising and finding the content poses more of a challenge. For example, with a visual lifelog captured using video content, at normal playback speed it will take a year to relive a year's lifelog [25], clearly it is necessary to develop organisation and search techniques that manage personal information archives and these need to be adequately evaluated. Replaying events from visual records of this form has found application in improving short term memory recall for memory impaired patients in clinical studies [24].

A more challenging and intriguing scenario is the use of PLs to construct a digital narrative describing incident(s) from the user's life in story-form. Since this is a topic generally poorly explored in information access, we outline its features in some detail in this paper. Stories, and their retelling, are fundamental to being human, in [31] Kearney declares that "stories are what makes our lives worth living". If we consider that PLs contain a person's life experiences captured through digital technologies, then the potential for telling personal and intimate stories from such collections is enormous [11]. A digital narrative in its simplest form is a series of causally related media artefacts that are brought together in some presentation mode to provide a meaningful and coherent dialogue. Brookes more clearly defines a model of digital narratives built from three separate functional layers, namely: the *structural*, *representational* and *presentational* [8]. Structural considerations are concerned with describing the narrative in simple abstract terms. The representation of the story seeks to capture the relationships between the various story elements in order to facilitate the reasoning required to build the end narrative. Finally, the presentation of a digital narrative applies the structural and representational knowl-

edge in order to choose and sequence story elements through reasoning and with a sense of aesthetic style.

One of the key issues in reviewing contents from a user's PL is the selection of material to be included in the presentation. Even for a visual diary simple rapid playback of all of passively captured images from during a single day rapidly becomes time consuming and tedious. This quickly becomes a much more significant problem when attempting to construct rich narratives taken from over an extended period of time. Overlapping with the issue of content selection, is the topic of presentation. Providing an effective interface to the data should enable the user to more rapidly and effectively access the presented information, and can potentially enable more information to be included in the output without loss of comprehensibility.

In our current work we are exploring the use of PLs for both search and narrative tasks. However, to understand our success and failure in this work, and to move our research forward, we need to establish measures by which our work can be assessed.

4 Evaluation Metrics

In order to evaluate information access applications for PLs experimental collections are required which are sufficiently large and diverse to represent the expected features of real user PLs. As a consequence, evaluations are wholly constrained by the number of participants who have collected data. Furthermore, there must be a sufficiently long period of collection to allow for an adequate number of personally significant and causally related events to have been captured. Issues of this sort are of course common for existing information access experimental work. Development of test sets inevitably leads to collections which have some limitations and are open to criticisms of the extent to which they are truly representative of real tasks. For example experimental results of retrieval evaluation tasks using crawled web search test collections are often cited as being too small to reliably reflect performance on the web itself which is several orders of magnitude larger. We explore our current work on the collection of large experimental PLs in section 6.

Assuming that suitable collections are available, performance in search tasks can be measured by a variety of metrics taken from traditional IR evaluation exercises, and personal digital narratives can take lessons from existing work in narrative studies and summarization.

For known-item ad hoc search with ranked retrieval one can apply the metrics of mean rank or mean reciprocal rank as used in existing workshop evaluation tracks, for example [26]. Where there is more than one potentially relevant item for the search query standard metrics of precision and recall can be applied. These have been applied in some of the limited existing studies in information access for PLs. For example, the Connections [37] system used precision and recall to measure performance [37], and also obtained anecdotal evidence of user satisfaction. Another metric explored in existing work is the analysis of the time taken by subjects to complete tasks using different approaches [32].

While quantitative methods have been used in some studies, questionnaires, surveys, observations of subjects searching behaviour and diary studies (the results of which can be used to note trends and peoples' behaviour) to establish user's searching, refinding and file sharing processes were the most common evaluation methods used by participants at the Personal Information Management (PIM) 2008 workshop [2].

The evaluation of narrative storytelling requires a little more extensive analysis. At present, there is no general consensus on how a narrative should be evaluated or the criteria for their evaluation. Researchers working in this area may fail to report their evaluation strategy with sufficient depth for it to be useful to others or repeatable [4], opt for a highly qualitative or informal approach to evaluation [23, 9] or omit evaluation (from their reports) altogether [33]. As outlined previously there are four major functions of our internal 'organic' stories which should be applied to their digital counterparts. These can be broadly condensed to two types of tasks. The first is the social sharing of stories which applies to both everyday and life stories. The second is personal review of stories for reminiscence or reflection, which is predominantly a function of our life story. As with IR strategies for PLs, the evaluation of reminiscence based tasks with personal narratives is problematic.

In order to practically and reliably investigate reflection such a task can only be evaluated with the participant who owns the data. The sharing of such stories is not bound by this limitation. Even if only a small number of participants collect PLs, if they are willing to contribute their material to a research community, shared narratives may be successfully evaluated by distributing them to a large number of external participants. .

One seemingly obvious approach to the quantitative evaluation of digital narrative output within PL content would be to adopt the approaches taken with the summarisation, as for example adopted for the evaluation of unstructured video Rushes Content in [29]. Within this TRECVid summarisation evaluation, participating sites seek to compress raw unedited video footage, with high numbers of retakes and large amount of consequent repetition, to within 2% of its original length using a variety of techniques. The resulting video content is then evaluated using a standard set of benchmark measures including coverage and flow or usability of the summary. Coverage of the narrative of the original content is a sensible metric to be included in

narrative evaluation, however, it should be noted that narrative is not just summarisation. In fact there are clear distinctions between the goals of the two. A good narrative is to be emotive, expressive and most importantly engaging, while a summary attempts to communicate as much information in as compact a space as possible. As such other measures must be considered to temper the notion of coverage and more fully consider the goals of a narrative. We believe that there are four major dimensions by which a narrative should be evaluated:

1. *Form and content*: The choices of what to show and how to show it are fundamental to a good digital narrative [38]. Within PL narratives multimodal content will typically be presented to an audience. As such both the fluency of the media chosen and the mode in which it is presented will have significant impact on the communication of the meaning and themes of the story. We must as such consider how to appropriately measure or evaluate and compare such factors.

2. *Coherence*: Coherence is integral to a good story: The quality of the end narrative is determined by how well organised and well paced the content is presented, and the rate at which the plot unfolds. Within interactive and digital narrative there is a careful balancing act to be maintained, between the story's coherence and the provision of agency to the audience as to how the story unfolds [10].

   Coherence can potentially be measured through the level of control provided, the perception of distortions or 'jumps' in the story, and/or attempting to ascertain the overall flow of the story.

3. *Expressiveness and affect*: Our life stories are experiential and are often highly bound to our emotional (or affective) state at that time. As such, the resulting stories must sufficiently express these affective and experiential components. There are many ways in which the affect may be represented within narratives, one method is the use of colour to represent mood or emotional state [38]. The evaluation of expressiveness should not focus on the mode of presentation however, but rather should consider how successfully the narrative conveys and communicates these experiential aspects. Approaches to ascertaining such metrics might be to employ heuristics such as those suggested in [14].

4. *Engagement, Impact and feedback*: Brooks asserts that an audience should actively desire to know the outcomes of a story, and with a better narrative, the experience becomes much more engaging achieving just that [8]. This engagement is a vital component for any story, and should ideally be reflected in PL-based stories, and as such we must carefully consider it within any evaluation. Although it is likely to be difficult to accurately and reliably elicit such metrics, we assert that evidence of personal engagement with the story should be recorded and reported in some manner be it qualitative or quantitative.

An alternative approach to evaluating narrative would be to adopt the rules which apply to more traditional or conversational discourse. For example, Grice [21] outlines a series of maxims under four categories: quantity, quality, relation and manner. These could easily be adapted to provide a suitable framework for the evaluation of many of the properties of a good narrative, primarily its form, plot, and discourse.

Brooks also notes that the better the representation, the more choices in the narrative paths for the story and this leads to a much better end narrative [8]. As such it may be necessary not only to evaluate the output and form of the end narrative, i.e. its presentation, but also to examine in some way the representation. Elson and McKeown [16] outline measures for the quality of a narrative's representation as: *expressiveness* (the range of stories it can encode), *robustness* (how well it handles partial or abstractly told stories), *formality* (how well its symbols lend themselves to formal analysis or automatic processing) and *usability* (how intuitive the representation is for collecting encodings from those not versed in narrative theory). The evaluation of the representational form of the narrative under such measures of quality could be included, however it is likely to be dependent on the task under evaluation and the goal of the investigation.

5 Previous Work in Information Access for PLs

Although the topic of collection materials for PLs and their utilization is still very much in a development stage, a number of relevant studies have already been reported in each of the areas discussed so far in this paper. This section summarises some of the most notable recent work in these areas.

5.1 Retrieval from PLs

With regard to retrieval tasks the following studies describe some of the most relevant work reported in this area to date. In [36] [37] PL retrieval algorithms are developed and evaluated on personal archives for 6 subjects consisting of computer activity over a period of 6 months. The subjects submitted 3-5 queries, which they appeared to freely generate from their memory. These were multi-item standard ad hoc type queries, e.g. "locate all items associated with writing of EVIA paper." To create oracle results for these queries the results of this query across different search

engines were pooled together, and the subject rated the relevance of the pooled result set. Other researchers, while limiting their evaluation to emails from personal archives, moved beyond this limited free recall approach to generate large numbers of tasks for known-item tasks. In [15], 30 queries with target email items which had been sent to a large number of people in a company were manually created. These queries were then entered by test subjects into the *Stuff I've Seen (SIS) interface*, resulting in the retrieval of various items from their personal collections including the target email. Subjects were then required to locate the target email, thus allowing for testing of various versions of the SIS interface. While this technique proved useful for testing features of PL interfaces, it would probably not be appropriate for generating tasks for all types of PL system experiments. For example, for evaluating PL retrieval algorithms where a user is required to use recalled content and context to form a query, i.e. if we generated tasks by giving a subject a task description of emails that had been sent to a group, the subject may have no recollection of these emails, and therefore could not recall content or context data with which to form a query.

In Elsweiler's work [17] a more personal approach to task generation is adopted. This work presented a framework for a task based approach to PL user evaluation. Specifically, over a three week period subjects recorded email tasks (task = email viewed and the purpose for which it was viewed, e.g. relocate email which contains Joe Blogg's phone number). They also allowed for the generation of additional tasks in a manner similar to Dumais's work [15] described in the previous paragraph. However they did not provide the query terms, instead providing a task description which simulates a 'real world' task the subject may engage in. Tasks were then categorized into three distinct types: tasks requiring a specific piece of information from within a computer item; tasks requiring a specific computer item; and tasks requiring information from multiple computer items. Tasks recorded by a subject were then presented to the subject for them to retrieve the task using a provided interface. This approach allowed for comparison of retrieval performance across different task types, and importantly the evaluation of a PL retrieval system, in a structured manner using the PL owners themselves. This approach could potentially be extended and used as a first step towards a much needed standard for evaluation in the PL domain. Elsweiler et al. examined this task generation approach only on emails and web pages, and thus its portability to other item types is not guaranteed.

By linking PL items, we believe textual descriptions from linked to items may be obtained which could prove beneficial in search, both for items of a non-textual nature such as photos or recordings of music, and items for which the user cannot recall the textual contents in sufficient detail. A number of existing studies have begun to explore the use of linking and annotating items. For example the PHLAT system allows individuals to add tags to their computer items [13]. However, in general the retrieval methods used for this data at present use existing backend algorithms and appear to operate on a simple boolean premise. For example [15] [20] where the queried context data is either present or not present in the database items. We believe that to truly harness the power of item linking and exploitation of related context data, items should be annotated with many rich sources of context data to support linking and search, and that novel retrieval algorithms tuned to PL tasks need to be developed which exploit these annotations.

5.1.1 Pilot Studies in PL Retrieval at DCU

We are interested in investigating individuals' memories of past interaction with items in their PL, the types of context data they recall, how recalled context data varies across different item types, and how individuals' memories change over time. In particular, it is our hypothesis that integrating remembered and partially remembered context information into a traditional IR algorithm will prove beneficial for PL retrieval.

In order to begin our investigation of these issues, we conducted a pilot study to test the types of context data recalled by an individual and to examine the utility of remembered context data in retrieval. A small PL was gathered for one individual over a 6 week period from the middle of July 2007 to the end of August 2007. To gather the PL, a logging application monitored activities carried out on the individual's computer during the period and additionally regularly prompted the user to manually log the location at which these happened. The retrieval experiment was a known-item search task where the user searched for a single partially remembered item from within the data collection. To generate the query tasks for the pilot study, following the 6 week collection period, the test participant identified a number of key events that occurred during the time span of the data collection. These events included: a friend's birthday, several meetings in the office, and dinner at a restaurant. Following this, the participant was required to recall the activities performed on their computer around or close to these key events. Each of the 27 test cases generated in this manner then consisted of a remembered computer file (e.g. email, word document, IM) and remembered context data associated with the file.

If PLs are to be recorded and accessed over an extended period of time, 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 detail soon after an event occurred, however with time memory fades and it is anticipated that less

will be remembered a significant amount of time after an event occurred. A further free recall study on the subject 6 months after the initial test case generation process was thus conducted to explore the effect on remembered content and context after this time interval. This study provided us with remembered content and context from the original 27 test cases at the 6 month interval.

The Mean Rank and the Mean Reciprocal Rank [26] of the target documents were calculated across the two query sets consisting of remembered content with various types of remembered context (e.g. query set 3 = content + location context + date context). The results of this study suggest that the types of context data an individual is most likely to recall in the long-term and that this context data is beneficial in PL retrieval. In particular, the results indicate that combinations of content and context can improve retrieval accuracy from an PL; and that context appears to become a more important factor over time as the subject's recall of contents declines. Full details are described in [18] [27]

5.2 Visual Diaries

Before examining relevant existing studies in narratives, we examine the simpler case of visual diaries which have so far been explored more extensively. Several studies have been conducted in recent years which passively capture a visual record of the wearer's daily activities. In this overview we focus on work using the SenseCam developed by Microsoft Research in Cambridge, UK [24]. The SenseCam is a small wearable device that passively captures a person's day-to-day activities as a series of photographs. It is typically worn around the neck, and so is oriented towards the majority of activities which the user is engaged in. Anything in the view of the wearer can be captured. The device incorporates on-board sensors to detect changes in light levels, motion and ambient temperature and then use these to determine when it is appropriate to take a photo. At a minimum the SenseCam will automatically take a new image approximately every 30 seconds, but sudden changes in the environment of the wearer, detected by the onboard sensors can trigger more frequent photo capture. The benefits of this include: the ability for a user to record events without having to sacrifice their participation, aiding memory and recall; and providing insight into a person's life and activities [24]. Wearing a SenseCam, the wearer can very quickly build a large archive of photos, or a visual lifelog. Within just one week over 20,000 images may be captured when wearing a SenseCam. When worn over a year the lifelog photoset can grow to over one million images.

At DCU we have been gathering archives of SenseCam images in various collection efforts. Multi-month visual lifelogs have been gathered by 5 researchers with another researcher gathering an archive of over two years data. This large collection has over 2.5 millions images. Based on our experiences of gathering this archive, we have found that the photos captured during this extended period of continuous lifelogging are vastly different in content to what one would expect to find with conventional digital photos. We have found that images often do not have salient objects, many of them are either low quality or useless (40%), and the types of scenes/objects captured differ greatly from conventional photo collections. This is because the SenseCam is capturing aspects of a wearer's life that are not normally photographed, for example 20% of the photos we observed contained the wearer's hands, 20% also show the LCD screen of the wearer's office computer screen or laptop, while only 15% of these are taken inside of the office of the wearer. The personal nature of the lifelog photo collections, when taken from the viewpoint of the wearer, provide intimate details of the wearer's work and social activities. Other observations support this point, such as the fact that 6% of the photos show people's faces (friends, work colleagues and family members), 4.5% show the details of meetings and 2% of the photos show the wearer reading.

Evaluation of SenseCam derived visual diaries carried out so far at DCU has focused on examining the nature of the visual lifelogs, to better understand what information is contained in them. In addition to this, we have carried out an evaluation of approaches to managing visual PLs, such as how best to organise them into a series of events, how best to chose a keyframe to identify an event, and how visual feature detection tools (such as faces, horizon, cup) perform on the SenseCam photos.

Research on the use of visual diaries to improve memory in clincial patients with cognitive dysfunction has used recall of autobiographical events as an evaluation criteria. This work compared recall using the SenseCam to recall using both a written diary and no memory aid [24]. Other experimentation has been ongoing to examine how effective SenseCams are as a memory aid, for example one study on user recall after wearing SenseCams investigated differentiating between different types of recall; did the user 'remember' the event, or did the user simply 'know' that the event existed? These experiments typically take place over a short period of time, for example a week [34]. A key question for our work is to explore the effectiveness of visual diaries as a memory aid both in itself and in terms of their impact on memory itself in terms of rehearsed recall over the longer term. Moving beyond simple diary PL collections, digital narratives provide a much richer framework for presentation of content from a PL.

5.3 Narratives

From the components of a PL, two major types of stories may be told. The first being a conversational, or 'everyday' story. Within our social interactions on a day-to-day basis we frequently share stories with our colleagues, friends and family. Often they are anecdotal and are used to recall a moment of humour or a recent event. As such their function is primarily routed in socialising and sharing of experience(s). The second type is the 'life story'. Life stories are internal ever evolving narratives which integrate significant or emotional events and update retrospectively to account for changes in our perception and remembrance of past events, values and beliefs [28].

Unlike everyday stories, the 'life story' provides us with the opportunity to examine a set of causally related events, framed from the perspective of our current self and within the context of our larger life experiences, motivations, goals and ambitions. It provides three main utilities to us: first it allows purposeful reflection on one's life events and their development in order to affect or determine future actions (directive); second it enables us to share aspects of our lives with others, often as a means to develop social bonds, intimacy and/or empathy (social); or finally, it affords a review of life experiences to search for meaning in them (self) [7].

Due to the recent interest in PLs provided by tools such as the SenseCam, the composition of 'everyday' stories from such archives has seen some limited exploration in a number of small scale studies. In [19] images from the SenseCam along with associated GPS location information was presented as a means to recount a 'trip-based' experience as a lightweight story in the form of an animated slideshow composed of SenseCam images. Harper et al. [23] conducted a study with six participants into user-created digital narratives composed from SenseCam captured images. The study is very much focused on participants' perceptions of a manually created end narrative, and yields some interesting outcomes particularly highlighting the usefulness of such images in reflection and reminiscing over life experiences. Of most relevance, but perhaps not wholly within the domain of PLs, is the work of Appan et al. [4]. This investigated the composition of digital narratives for everyday experiences using media such as photos, gathered during the user's day-to-day activities. While there has been some exploration of conversational stories, we are not aware of any work exploring digital equivalents to the life story. Given the volume, complexity and richness contained within PLs there is great potential both for the elicitation of such stories and the evaluation of them.

6 PL Activities at DCU

In order to move on from our initial work on PL access applications, our group at DCU are currently engaged in collection of large experimental PLs. 4 subjects are currently actively engaged in collecting personal data to form a PL. Ideally these collections would be developed over very extended periods, possibly running into many years. For practical reasons, i.e. we cannot wait years to continue our research, we are currently limiting the next phase of the collection period to one year. While not perhaps ideal, we should emphasise that the collections resulting from our current work will be much larger and more diverse than other collections made by ourselves or others to date.

The following timestamped data is being collected for each of our subjects over the one year collection period:

- Home and Work Computer Activity: every item (email, word document, web page, etc) accessed by the user, with time of access, contents of item and path to item information is being recorded. To do this a program which combines the output of Digital Memories software [1] and Slife software [35] running on the individual's computers was developed.
- Mobile phone activity in the form of call logs and SMSs are recorded. An application was readily available to capture mobile phone call logs, and an application was developed to capture sent and received SMSs.
- GPS data, from which location names, light status and weather conditions can be derived, and Bluetooth data, from which people present can be detected, are also captured on the mobile phone, using software available within our group.
- SenseCam images and digital photographs are also stored.

Physiological and heart rate data is also being captured to monitor physiological condition and infer emotional state. It was decided that it would not be practical to capture this biometric data over the course of the entire year, since the sensors are quite physically intrusive to daily life, so this data is being collected for a one month period by the 4 subjects.

The 4 PLs generated over the course of the year will allow us carry out detailed investigations with what approach 'real' PLs. However, it is acknowledged that solid experimental results necessitate a larger number of test subjects. Thus to provide further support for our experimental results, more subjects will be obtained during the course of the year for smaller time frames to capture their own PLs.

In order to collect their PL, each of the participants has been provided the following apparatus.

Hardware

- Nokia N95 mobile phone: The N95 is a sophisticated mobile device which runs the Series 60 platform enabling mobile applications to be developed and deployed on it. It also includes a built-in 5-megapixel camera and large storage capabilities (up to 8GBs) making it ideal for use in the long-term collection of data.
- Microsoft Research SenseCam: Although not commercially available, we are fortunate to have been provided with a number of these devices by Microsoft Research.
- Garmin GPS Loggers: These small portable devices automatically regularly log the participant's location. Due to the limitations of GPS technology, typically these are only used when the participant is outdoors and on the move.

Software

- Desktop Activity Recording Software: Users are asked to run desktop activity logging software constantly on both their home and work computers. For those running Windows-based machines, they are provided with both the Microsoft MyLifeBits [1] and SLife [35] software, while those running Macintosh OSX are only required to run SLife. These software suites record all application use and content access on a desktop or laptop computer.
- Mobile Context Recording Software: The *Campaignr* software, provided to us by UCLA (USA), has been deployed on the N95 devices. Participants run this software constantly to capture contextual factors such as their location through Wireless, GSM and GPS sniffing and the persons in proximity to them through Bluetooth device sniffing.
- Mobile Activity Recording Software: A proprietary piece of software enables the logging of missed calls, made and received calls and sent and received text (SMS) messages on the N95.

These various data sources will be integrated into a PL for each user for investigation of PL information access applications. Part of the user studies will also focus on determining the accuracy of our context derivation methods by comparing derived context data to that experienced by test subjects. Participants will also complete a number of questionnaires over the course of the PL build up year to investigate various factors of PLs, such as items the participants consider important within their collection, items that they believe are likely to want to retrieve in the future, and the types of content and context data subjects recall about items of different type and importance level.

6.1 Structuring the Data

An important aspect of PL collections is how to structure the data for effective information access. For example, the many thousands of SenseCam images taken as ongoing sequences should be segmented into meaningful units for browsing and retrieval purposes. In the domain of digital video search, there is a well known technique called shot boundary segmentation that identifies the camera recording boundaries that compose a piece of edited video. Often the video clips between the shot boundaries (video shots) are used as the unit of retrieval. Our group have carried out extensive evaluations of event segmentation for SenseCam collections to identify the best approaches to integrating context and content data to achieve high precision event segmentation [22]. Following on from the event segmentation process we have been exploring the judicious selection of a key-image or other surrogate data to represent an event for the end user [6]. In our work we have focused on evaluating techniques to represent each event by a single image or keyframe. In addition to event segmentation and representation activities, we are exploring how successfully conventional digital image concept detection techniques can be applied to visual lifelogs. Finally we are also planning to evaluate a procedure for automatic annotation of people present at the time of capturing data for a PL, by evaluating how to effectively identify friends, family and work colleagues by means of Bluetooth device logging [3].

Beyond the structuring of the SenseCam data, we are beginning to investigate how the other sources of data within a PL can be linked in support of information access applications.

6.2 Search

Using our new PL collections we plan to conduct extended versions of the experiments reported in [18] [27]. We also plan to use other task generation processes described in [17], and the use of information contained in digital schedulers to help in the task generation process.

In addition to this, we are also seeking to better understand circumstances in which users will remember different features associated with personal information and the information itself, and how they will subsequently describe it. As part of this investigation we are looking at models of the underlying cognitive processes in human information storage and retrieval [12]. In order to do this we are conducting a range

of user experiments via questionnaires and interviews with our collection participants and others. Although of course as mentioned earlier, care needs to be exercised in doing this not to modify the typicality of the collections by requiring users to rehearse recall of items from their PL, with the consequence of affecting the reliability of results of subsequent retrieval experiments.

6.3 Digital Narratives

With regard to digital narratives, we are just beginning to explore the possibilities within such collections. We hypothesize that the richness of content and context captured in a PL will enable the approximation of our internal 'organic' conversational and life stories with digital equivalents. Additionally, we believe that such digital equivalents can also serve to support, prompt and assist the the functions of our internal stories as outlined previously. The investigation will be carried out as follows.

First a preliminary study will seek to ascertain how personal life stories should be represented through digital narrative. To determine this we have planned a qualitative probe using a large number of participants. Participants will be asked to construct a number of simple conversational stories and one more extensive life story using media artifacts already available to them, including photos and videos or text from emails for example, or relevant content that can be easily found online on sites such as Flickr. The stories created will then be probed during one-on-one follow-up interviews, the results of which will be coded to determine if there are commonalities in the approaches to representing the story which could be developed into a model. The conversational stories created will also be used to probe the fluency of various media types and their affect on the communication of the story. These will then be shown to new participants who will be asked to quantitatively analyse the communication of the story. To measure this we will develop an evaluation framework based on the dimensions mentioned in the previous section. This framework will probably take the form of a questionnaire that will probe the participants' reactions to and perceptions of the story presented to them.

Utilizing the knowledge gained from these studies, we will next build a semi-automatic narrative generation engine as described in [11]. The structure of the narratives will be event-oriented as suggested by the related work in [4]. In the authoring process of this system, users will browse the archive in an event-oriented manner. They will locate and select the episodes which are relevant to the story, marking them for inclusion. After which, they will be asked to make basic plot and aesthetic choices to determine the presentation of the narrative. Following this the system will enter a generation process in which the selected content and author choices are constructed into a coherent story.

With the system built, the final activity in our study will be to validate the output of the system. Example output will be used to evaluate the hypothesis that digital life stories can be used to enable personal reflection and sharing of meaningful life experiences.

7 Conclusions

In this paper we have introduced the emerging area of information access for personal lifelogs. We have outlined the differences between PLs and more conventional existing document collections and envisaged information access tasks for them. We have also summarised work to date by ourselves and others exploring and evaluating information access for PLs. In particular we highlighted their use in personal search, browsing and narrative. The paper concludes with a description of our current work in collection of large PL test set for use in our ongoing experimental work.

Acknowledgments

This work is funded by grant CMS023 under the Science Foundation Ireland Research Frontiers Programme 2006 and a postgraduate studentship award from the Irish Research Council for Science, Engineering and Technology, and is supported by Microsoft Research under grant 2007-056.

References

[1] MSR MyLifeBits Project - `http://research.microsoft.com/barc/mediapresence/MyLifeBits.aspx`.

[2] *PIM 2008: Proceedings of Personal Information Management, Workshop at CHI 2008*, Florence, Italy, 2008.

[3] K. Aizawa, D. Tancharoen, S. Kawasaki, and T. Yamasaki. Efficient Retrieval of Life Log Based on Context and Content. In *The 1st ACM Workshop on Capture, Archival and Retrieval of Personal Experiences (CARPE 2004)*, pages 22–31. New York, New York, USA, October 2004.

[4] P. Appan, H. Sundaram, and D. Birchfield. Communicating Everyday Experiences. In *SRMC '04: Proceedings of the 1st ACM workshop on Story representation, mechanism and context*, pages 17–24. ACM, 2004.

[5] G. Bell and J. Gemmell. A Digital Life. *Scientific American*, March 2007.

[6] M. Blighe, A. Doherty, A. F. Smeaton, and N. E. O'Connor. Keyframe Detection in Visual Lifelogs. In *ETRA 2008 - 1st International Conference on Pervasive Technologies Related to Assistive Environments*, 2008.

[7] S. Bluck and T. Habermas. The Life Story Schema. *Motivation and Emotion*, 24(2):121–147, 2000.

[8] K. M. Brooks. Do story agents use rocking chairs? The theory and implementation of one model for computational narrative. In *MULTIMEDIA '96: Proceedings of the fourth ACM international conference on Multimedia*, pages 317–328, New York, NY, USA, 1996. ACM.

[9] K. M. Brooks. Metaliner cinematic narrative : theory, process, and tool. In *Thesis (Ph.D.)*. MIT, 1999.

[10] A. Bruckman. The Combinatorics of Storytelling: Mystery Train Interactive. *Unpublished paper*, 1990.

[11] D. Byrne and G. J. F. Jones. Towards Computational Autobiographical Narratives through Human Digital Memories. In *SRMC 2008 - 2nd ACM Workshop on Story Representation, Mechanism and Context*, 2008.

[12] Y. Chen and G. J. F. Jones. Integrating Memory Context into Personal Information Re-finding. In *Second Symposium on Future Directions in Information Access (FDIA 2008)*, 2008.

[13] E. Cutrell, D. C. Robbins, S. T. Dumais, and R. Sarin. Fast, Flexible Filtering with Phlat - Personal Search and Organization Made Easy. In *CHI 2006: Conference companion on Human factors in computing systems*, pages 261–270, New York, NY, USA, April 2006. Montreal, Quebec, Canada, ACM Press.

[14] E. de Lera and M. Garreta-Domingo. Ten Emotion Heuristics: Guidelines for Assessing the User's Affective Dimension Easily and Cost-Effectively. In *Proceedings of the HCI07 Conference on People and Computers XXI*, page 40, 2007.

[15] S. Dumais, E. Cutrell, J. Cadiz, G. Jancke, R. Sarin, and D. C. 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.

[16] D. K. Elson and K. R. McKeown. A Platform for Symbolically Encoding Human Narratives. In *Proceedings of the AAAI Fall Symposium on Intelligent Narrative Technologies*, pages 29–36. AAAI Press, 2007.

[17] 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.

[18] M. Fuller, L. Kelly, and G. J. F. Jones. Applying Contextual Memory Cues for Retrieval from Personal Information Archives. In *PIM 2008: Proceedings of Personal Information Management, Workshop at CHI 2008*, 2008.

[19] J. Gemmell, A. Aris, and R. Lueder. Telling Stories with MyLifeBits. *Multimedia and Expo, 2005. ICME 2005. IEEE International Conference on*, pages 1536–1539, 2005.

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

[21] H. P. Grice. Logic and Conversation. In P. Cole and J. L. Morgan, editors, *Syntax and Semantics: Vol. 3: Speech Acts*, pages 41–58. Academic Press, San Diego, CA, 1975.

[22] C. Gurrin, A. F. Smeaton, D. Byrne, N. O'Hare, G. J. F. Jones, and N. E. O'Connor. An Examination of a Large Visual Lifelog. In *AIRS 2008 - Asia Information Retrieval Symposium*, 2008.

[23] R. Harper, D. Randall, N. Smyth, C. Evans, L. Heledd, and R. Moore. Thanks for the Memory. In *Proceedings of HCI 2007, The 21st British HCI Group Annual Conference*, page 10. BCS, 2007.

[24] S. Hodges, L. Williams, E. Berry, S. Izadi, J. Srinivasan, A. Butler, G. Smyth, N. Kapur, and K. Wood. SenseCam: A Retrospective Memory Aid. In *UbiComp 8th International Conference on Ubiquitous Computing*, 2006.

[25] T. Hori and K. Aizawa. Context-based Video Retrieval System for the Life-Log Applications. In *The 5th ACM SIGMM International Workshop on Multimedia Information Retrieval, MIR 2003*, pages 31–38. Berkeley, California, USA, 2003.

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

[27] L. Kelly, Y. Chen, M. Fuller, and G. J. F. Jones. A Study of Remembered Context for Information Access from Personal Digital Archives. In *IIiX 2008 - 2nd International Symposium on Information Interaction in Context*, 2008.

[28] D. P. McAdams. The Psychology of Life Stories. *Review of General Psychology*, 5(2):100–122, 2001.

[29] P. Over, A. F. Smeaton, and P. Kelly. The TRECVid 2007 BBC Rushes Summarization Evaluation Pilot. In *TVS '07: Proceedings of the international workshop on TRECVID video summarization*, pages 1–15, New York, NY, USA, 2007. ACM Press.

[30] L. Page, S. Brin, R. Motwani, and T. Winograd. The PageRank Citation Ranking: Bringing Order to the Web. Technical report, January 1998.

[31] R. Kearney. *On Stories*. Routledge, London/New York, 2002.

[32] M. Ringel, E. Cutrell, S. Dumais, and E. Horvitz. Milestones in Time: The Value of Landmarks in Retrieving Information from Personal Stores. In *Interact 2003*, 2003.

[33] W. Sack. Stories and Social Networks. In *AAAI 1999 Fall Symposium on Narrative Intelligence*, Menlo Park, California, 1999. AAAI Press.

[34] A. Sellem, A. Fogg, S. Hodges, and K. Wood. Do lifelogging technologies support memory for the past? an experimental study using sensecam. In *Conference on Human Factors in Computing Systems, CHI '07*, 2007.

[35] Slife Labs - http://www.slifelabs.com.

[36] C. A. N. Soules. *Using Context to Assist in Personal File Retrieval*. PhD thesis, School of Computer Science, Carnegie Mellon University, Pittsburgh, PA, USA, 2006.

[37] C. A. N. Soules and G. R. Ganger. Connections: Using Context to Enhance File Search. In *20th ACM Symposium on Operating Systems Principles (SOSP'05)*, pages 119–132. Brighton, United Kingdom, 2005.

[38] J. Tanenbaum and A. Tomizu. Affective Interaction Design and Narrative Presentation. In *AAAI 2007 Fall Symposium on Intelligent Narrative Technologies*. AAAI Press, 2007.