Are episodic context features helpful for refinding tasks? Lessons learnt from a case study with lifelogs

Yi Chen
CNGL, School of Computing
Dublin City University, Dublin 9, Ireland
chenyidcu@gmail.com

Gareth J. F. Jones
CNGL, School of Computing
Dublin City University, Dublin 9, Ireland
gjones@computing.dcu.ie

ABSTRACT

Both psychological theories and findings in information science suggest that people may remember the episodic context of previously encountered information. This implies that a user's episodic memory might be utilized to improve the efficiency or effectiveness of refinding tasks. In this paper, we report a case study which aims to explore the feasibility of integrating episodic context into the design of information refinding systems. The subjects in this study collected 20 months of rich contextual data along including the full text of all documents, emails, web pages and so on, which they accessed during the collection period. We developed a "memory-friendly" system based on psychological theories to test the hypothesis through user studies requiring the subjects to find their personal data using this system. From examination of the user activity log and a post-task questionnaire, we found that although our designed features, which support or utilize episodic context or autobiographical memory, were not used as frequently as we expected, they did improve the effectiveness of the refinding tasks.

Keywords
Refinding, memory, episodic context, lifelog, case study.

1. INTRODUCTION

The expansion in the availability of online information resources and development of search technologies has made it easier for people to look up information, e.g. from the World Wide Web, to add to their knowledge or data collection. While there has been much focus on them carrying out new ad hoc searches, people often need to find information that they have found or encountered before, referred to as *refinding* or *personal information search*. *Personal information* includes any information that belongs to a person or is related to their past experiences, including all the information that the individual keeps, receives, experienced, or is about them [1]. Since the information in a personal information space is highly related to their past experiences or knowledge, refinding tasks usually require them to explicitly recall certain information related to the target, e.g. file name or location, in order to proceed. Currently, people can refind known items using keywords or target attributes, which they remember, with advanced desktop search tools such as spotlight, Google desktop or windows search. Alternatively, they can navigate to the parent directories in a specific application (e.g. mail client, "Finder" or "explorer"), given that they know where the items are stored. We believe that it is important to adopt a **memory-oriented** and user-centered approach to design future systems for people to easily re-access their personal information (without intense effort in managing and remembering corresponding information). In fact, several researchers have also argued for the importance of supporting a user's memory in refinding tasks (e.g. [2-6]), and designed systems to cater for the user's memory. For example, Elesweiler et. al. [6] built a personal photo refinding system based on their study of people's daily memory problems.

Our study is inspired by cognitive psychology theories of memory, in particular, the associated memory network, as well as findings from some existing studies (e.g. [7-9]**),** which suggest that people may have better memory of events or temporal spatial context related to a document that they have accessed than the exact details of the documents, such as the exact file name or keywords associated with it. For example, in the study by Gonçalves et al. [7], their participants reported rich contextual information in addition to attributes of personal documents. Ringel et al [9] found that people sometimes remember a rough temporal relation between their activity of accessing a digital item and their personal or important public events, and that these events can help people to more efficiently find digital items. These findings are supported by theories of human memory. In the next section, we review psychological theories which provide evidence to support these findings, and further discuss how cognitive models and theories apply to the design of personal information refinding systems.

ACM 978-1-4503-2976-7/14/08, $15.00.
http://dx.doi.org/10.1145/2637002.2637013

2. MEMORY IN REFINDING

2.1 Associative Memory Network

Memory is a mental system that encodes, stores and utilizes the information that a person encounters and experiences. The information in memory is not encoded and retrieved as a whole, but stored as and reconstructed (when retrieving) from small pieces of memories (called *engram* or *memory traces*) in an associative network, in which, memory traces are linked from one to another by various relations, e.g. belonging to the category, part of, accompanied, and following, etc. When one node (a memory trace) in the memory network is activated, it spreads the activation to its linked memory traces and so on. The stronger the association (links) between two memory traces, the faster the activation spreads and the greater activation it receives (see review [11] for more detail). The stimulus that triggers a memory trace is called a *cue*. So, how does this memory network affect a refinding task?

2.1.1 Memory in Search

Information refinding tasks involve cued-recall tasks, in which people are prompted to retrieve memory traces with the help of cues. When someone wants to find a piece of encountered information, he/she usually already has some idea about the target item, which means, some memory traces which link to the known aspects of the target item (as given by the task) should have been activated. This means that any memory trace may have a chance to be recalled, but only if it is directly or indirectly associated with the target item. For two memory traces to be associated, they need to have been encoded (presented in the working memory) together [10]. This means that things that are presented at the same time or one after another in adjacent time are likely to be associated, if they were both perceived by the person. For this reason, information about the context (location, weather, background scene), in which some information is encountered, is usually associated with the memory of the information, and is likely to be recalled.

2.1.2 *Contextual Cues in Refinding*

On the other hand, if the task context has evoked memory traces that are associated with the memory of a target item, more details of the target item can be recollected. According to *Adaptive Control of Thought* (ACT) theory [12], at the time of retrieval, the more activated nodes that link to a memory trace, the greater activation the memory trace can get, and the greater the chance that it can be triggered and pop into conscious awareness (and be retrieved). For this reason, a person is more likely to have better recall when he or she is in the same or a very similar context to that when they previously encountered the information, since many elements in the context can act as cues to activate corresponding memory traces (of the environment) which link to it; this is referred to as *environmental context-dependant memory* [13]. While we can hardly restore the physical environment to its state when the target items was encountered previously, *we can provide digital copies of these cues such as photos or names of location or people which may trigger the user's memory of elements in the previous context, which may further trigger memory of information encountered in that context*. Of course, not all digital records of the presented physical context will be good cues, since they are influenced by many factors, such as emotion, time lapse, and salience.

2.2 Autobiographical Memory and Directory-Based Navigation

Folder-based navigation is another method people often employ to find their documents or emails. In this sub section, we explore how to cater for memory features that are involved in refinding tasks using a navigation approach.

Problems with the traditional folder based navigation approach: When people remember the approximate storage location where the target information has previously been encountered, they can take a traditional location-based navigation approach to find the target item if its location remains unchanged. However, the rapid increase in the amount and type of items in personal information space means that it is becoming more difficult for people to remember the location of individual items, and, in fact, many items are stored automatically without ever having been explicitly organized to specific folders by the user. On the other hand, it is sometimes desirable that items are automatically assigned to directories in which the user will tend to "know" an item will be stored into, according to some simple schema. For example, if items are automatically sorted by item type (which users tend to know), e.g. emails, word, pdf, they can expect to find their target item in a corresponding folder based on the item type. Of course, there can be thousands of items that belong to the same type, which can make individual items difficult to locate.

Solution: Since people may remember the events or context in which information was encountered, we suggest another schema for organizing personal information, that is, organizing items according to the context or episodes. If the episodes are short enough, e.g. from a few minutes to a few hours, there are unlikely to be thousands of items to locate the target. Therefore, once the correct episode has been located, one should easily find the target items through browsing. Of course, the episodes can be further organized into larger units. To do this, we believe that the structure of human autobiographical memory may form a good model.

Autobiographical memory (AM) is memory about experiences and facts of one's self. According to Conway [15], there are three levels of events with different levels of specificity of knowledge that structures autobiographical memory: i) life time periods (e.g. when working at company X, or when living in city Y), ii) general events (e.g. holiday in France, working on a journal paper), and iii) event-specific knowledge. *Event-specific knowledge* contains a large portion of sensor-perceptual information of a single short period lasting from minutes to hours. Suppose that the "episodes" which we discussed in the previous paragraph can be modelled as *event-specific knowledge*. It should not only be time constrained to units of several minutes or hours, but also thematically distinct. The thematic features of each episode can be used to represent it, so that a user can easily be reminded of the content of the episode, and judge its relevance without seeing all the details. Once items are grouped by episodes, they can further be organized by general events, or even lifetime events. Consequently users can locate items in episodes by recognition of corresponding general events.

Time of events in autobiographical memory: Episodes or general events may also be sorted chronologically. Although we do not tend to remember exact calendar dates of activities or events that are not highly related to calendar dates, we tend to know the approximate temporal distance from the current time (now) or from a landmark event, or in a temporal scheme [16]. In fact, several systems have adopted the autobiographical memory model in designing information refinding applications. In particular, using landmark events to help users to locate their computer activities based on time (e.g. [9][17]).

2.3 Suggested Functions to Support and Utilize The User's Memory in Refinding

Based on the above understanding of human memory, we hypothesize that the following features may improve the efficiency of refinding tasks:

a) People are more likely to successfully retrieve a target from an information retrieval system if they are allowed to query using information from physical context.

b) If the system dynamically groups the data into general events and episodes, users are likely to know where to locate their target item or information through navigation.

c) Representative features for general events or episodes can help people to recall and locate potential targets contained within a given collection.

To summarize, the psychology literature suggests a promising future for integrating physical context and autobiographical memory in designing refinding systems. Therefore, we hypothesize that *if a personal information finding system can* *utilize a user's episodic memory of events and context related to the refinding targets, refinding tasks may be more effective and efficient*. In this paper, we report a case study that we conducted to investigate this hypothesis.

3. STUDY SET UP AND CHALLENGES

This section presents an introduction to the set up of our case study including the subjects, the data collection, and the methodology.

3.1 Overview

To test our hypotheses, we need to build a working system that incorporates the proposed features that utilize the user's memory of physical or episodic context. We believe personal information is highly interconnected, and that many tasks require more than a single piece or type of item or information. Therefore, we adopted an all-in-one design, like many desktop search systems (e.g. Google desktop, Spotlight, or Stuff I've Seen [3]), which manages all types of available personal information in one system. The design of the memory-supporting features is based on the psychology literature. Based on our hypotheses in section 2.3, the prototype system includes the following features/functions:

a) *Search:* In addition to traditional search fields such as keywords, subject or type, it should provide additional search fields relating to the physical context. We investigated which search options to provide in a pre-development study.

b) Navigation: It should dynamically organize information following the autobiographical memory model

c) *Preview/Faceted browsing:* The testing system should include a preview panel which presents thematic information or item of current directory. These items also act as facets to provide a faceted browsing function to better mimic the autobiographical memory model.

d) Landmark annotated timeline: When the items are sorted chronically, the system should provide the users with landmark mark events to help them locate items by time.

Of course, to evaluate the system, we need human subjects and their own datasets, which contain data to realize these functions, and links to their own memory. The difference from previous systems, which mainly rely on computer items and user generated data such as news, schedules and digital photos (e.g. [9]), is that we aim to utilize ubiquitous computing technologies for a wider range of contextual information relating to physical context.

Our study is enabled by unique sets of rich episodic contextual data together with recordings of corresponding information access on the user's computers, collected by three subjects over a period of 20 months from 2008 to 2009. The study reported in this paper is part of a larger project, exploring not only management, but also applications of such data. We refer to these collections of data as personal lifelogs (PLLs).

3.1.1 Subjects

The three subjects, referred to here as l*ifeloggers,* were research students who did their PhD research in topics related to lifelogging. Due to the value and uniqueness of their personal lifelog (PLL) collections, they were used by each of the students to carry out the main studies of their PhD research. Of course, due to privacy issues, the data was not made directly available to the other lifeloggers, but processed by algorithms and tools from the other two researchers, and was only exposed to the data owner him or herself to conduct experiments on his/her own computers. Notably, one of the lifeloggers was the first author of this paper (lifelogger C). All three subjects used a Nokia N95 as their main mobile phone during the collection period.

3.1.2 Data Collection

The lifelog data was captured using technologies available to the lifeloggers at beginning of the collection period in 2008. The collections contain the following data types:

1) Computer activities: every time a window came to the foreground, it was defined as an instance of computer activity. The attributes, timestamps, as well as the full text of the opened document, web page, email and so on were recorded.

2) Mobile phone activities: including the receiver and sender number or contact name of phone calls, short text messages (SMS), and full text of SMS messages.

3) Photos: images automatically taken using a Microsoft SenseCam[1] and digital photos actively taken with a digital camera are both included in our lifelog collections.

4) Geo-location: the lifelogger's location was captured by the embedded A-GPS on Nokia N95 mobile phones.

5) Bluetooth: the name of surrounding Bluetooth devices were captured with the expectation to represent the device wearer.

6) Biometrics: heart rate and Galvanic Skin Response (GSR) were captured for a one-month of the collection period. He limited period was due to the physical burden of wearing these devices.

7) Tweets: we also tried to collect Twitter posts at a later stage as the collecting progressed. Unfortunately, our subjects posted too few tweets to make analysis worthwhile

3.2 Challenges and Methodology

While these collections enable us to test the hypothesis with a working search system and data, the small number of subjects (three) makes it difficult to generalize our findings. To maximize the external validity, and minimize the issues of internal validity and reliability caused by personal difference in this study, we tried to collect exploratory data from as many subjects (in additional to the above subjects) as possible, and tested the components on other subjects whenever no long-term lifelog data was required. Our study is structured as follows:

Table 1. Participants in each study

Questions explored	Number of participants	
	lifeloggers	Other subjects
Types of likely remembered features	1	645
Validation of types of remembered features	3	0
Cue items for events	3	3
User interface unit test for layout design	1	4
Final systematic evaluation	3	0

In this paper, we focus on reporting findings from the final evaluation, but also review the other studies which led to the development of our system, by clarifying some practical questions, e.g. what exact types and forms of contextual attributes should be included in search, and how to automatically extract good contextual cues from a PLL based on psychology theories.

[1] http://research.microsoft.com/en-us/um/cambridge/projects/sensecam/

The rest of this paper is structured as follow: Section 3 discusses how a user's memory is involved in refinding tasks and introduces guidelines for functions which we hypothesize can support users in refinding tasks. Section 4 reports two of our empirical studies which investigate the exact attributes and parameters to be used in our prototype system, which is introduced in Section 5. In Section 6, we report the final evaluation of the system and our hypothesis, to determine whether these features can improve the effectiveness and efficiency of a personal information refinding system. The findings and issues raised are discussed in section 7.

4. PRE-DEVELOPMENT STUDIES

While the psychology literature explains why some information tends to be better remembered, and the mechanisms that enable a memory trace to be retrieved, it provides little to answer some important questions with regard to the development a memory-friendly refinding system. These unanswered questions include:

- What exact types of information tend to be well remembered for search targets?
- What data in PLLs can act as good cues to trigger people's memory of relevant events if we organize the data according to the structure of human autobiographical memory?

To answer these questions, we decided to seek answers through empirical investigations. In this section, we describe two series of studies which explored the above questions respectively.

4.1 Study 1: What Episodic Information Do People Remember?

To allow users to search by what they remember, and embed this most likely to be remembered information into search options in an information refinding system, we conducted a series of studies to explore the types of information people tend to remember, in particular, episodic contextual information related to a target This includes exploratory studies of remembered features collected from non-lifeloggers, and testing the reliability of these types of remembered features for the lifeloggers participating in the main evaluation study. (Details of this study can be found at [19]).

4.1.1 Maximally exploring types of remembered information

The first study combined a diary study and an online-survey to maximally explore the potential types of likely remember about reencountered information [19]. The findings are supportive of our hypothesis that people tend to remember physical context of previous encounters with information. Congruent with our discussion in section 2.1.1, most of the subjects reported that they remember one or more types of context, such as personal location, people nearby, weather conditions, light status and their emotional state. Of course, they also remember the content of the target, e.g., a summary of the content (of the story, article or movie), and part of the content (e.g. part of the lyrics, script or words), keywords or sentences in textual target, the name of website, title of the article or subject of the email. Visual features such as layout, background colour or salient visual elements were reported to be a remembered feature for a quarter of the tasks. Other types of information that people remembered about electronic items were: summary, gist of the meaning or other details (not exact words) of some content within or surrounding the target, such as descriptions and other comments on the page of an online shopping item, function of the website or application, and self-created content in it, e.g., "my comments" on an article. Since there may also be other types of refinding tasks in addition to what has been concluded in previous studies (e.g., [18]), and to make the study as generalizable as possible, we also designed questions to explore the types of potential types of tasks and targets. For example, we give a broad definition for refinding tasks ("looking for any information or item that you have encountered previously"), and provide a variety of examples of refinding tasks. As a result, in addition to lookup tasks and single item (known-item) search tasks, we added two other categories of refinding tasks: exploratory tasks and navigation tasks.

4.1.2 Validation of "Likely" Recalled Information

Unlike most other studies which also extensively examiner likely remembered types of information for refinding tasks (e.g.[7]), we also validated the recalled content. This is realized through another studies with our three lifeloggers.

In the study, each subject listed 100 targets, and purely based on their memory, entered details of each potential search field for the target such as keywords, city in which or people nearby when the information was encountered. After this, the information was inserted into an in-house implemented search engine and a list of potentially relevant items were retrieved. The subjects were then required to mark the relevant items. For each of the relevant items, attributes and metadata of the context for all occasions of accessing the relevant items were extracted from the subject's lifelog database, including: *keywords, extension (type of target item), date of visiting, month, season, day of week, part of week (week end, weekday), time range (e.g. 8am-9pm), people present, geo-location (e.g. city, street name), weather, file path, country, file name, from contact, to contact, device, year.*

Data Analysis*:* The above ground truth data were compared with what the subjects recalled. We then calculated the *hit rate* (the percentage of relevant items for each task which matches the recalled attribute or metadata) and *false query* rate (percentage of terms or values for a field that the subjects recalled that have no matching items in the relevant results). This enabled us to measure the reliability of recall and usefulness of a field:

Reliability of recall= *hit rate* x (1– *false query rate*)

Usefulness = *hit rate* x (1– *false query rate*) x *frequency of recall*

We found that: the extension (item type), country of the person, name of the contact who sent the email or SMS and part of week, are the most reliably remembered features. Attributes or features which seem to be most useful (frequently and reliably recalled) include: extension, keywords, country, geo-location, month, season, and part of week.

In the design of our search system, while all likely recalled information are allocated a search field if the information is available in our prototype lifelogs, the reliable items are given higher weight in the IR system, or treated as filters.

4.2 Study 2: What Information can Act as Good Memory Cues to Represent Events?

The second series of pre-development studies concerned automatically extracting contextual cues from lifelogs.

4.2.1 Methods

We adopted an experimentation approach which sampled potential cue items strategically, and examined their utility as memory cues for episodes and general events. Machine learning algorithms were trained using this data and tested through another user experiment. To strategically generate the training data set, we selected the types of potential cues and factors that influence the strength of the cues based on psychology literature (section 3.2) and two preliminary studies.

The first study [21] invited three temporary lifeloggers who only collected 1 to 2 days of lifelog data (mainly activities on their computers captured using software called "timestamp"). We employed a card sorting approach and cued-recall tests to investigate the types of data which are most remindful of computer activities. A rich variety of items were found to be good memory cues for computer-centred activities (episodes which focus on interacting with computers), including: name of the desktop applications used, the name of websites visited, desktop screenshots, information which represented the main content of computer activities, and the subject or the contact of email, etc. We also found that the subjects had better memory of the day on which they collected the reviewed their short-term lifelog data, than others day before and after that day.

The other study adopted a self-experimentation approach. It was designed to obtain better insights about factors in addition to types of items that make a good memory cue. For example, the self-experimentation found that manually taken digital photos tend to be better cues than automatically captured photos, and with reflection on the data, the subject and examiner proposed a list of hypotheses for the factors which may bring about such differences in the strength of cues, e.g. the image quality, importance of events in which the images were taken, the number of visits to the images; distinctiveness and density are two factor that determine the strength of facets type cues.

The training data set for our cue strength estimating algorithm was initiated by sampling each type of potential cue item from each level of each factor. The cue strength is annotated by the data and memory owners themselves through a recall experiment. For each of the cue items presented, we asked the subjects whether they recognize the specific event or general event. This was followed by some other rating questions regarding their memory of the event if they recognized it, e.g. the month during which the event took place. The measure of remindfulness, which we call a *cue strength score*, was calculated from multiple ratings for each "cue item". The sampling of the cue items was done strategically, so that the data covered the full range of values for each factor which we hypothesized to be influential to the strength of the cues. The values for each of the hypothesized factors were extracted from the lifelog database as potential predictors for each rated cue item. Regression models were used to train algorithms to predict the "cue strength" of cue items for different types of events, namely: episodes (specific events), general events, and landmark events (which are events which can act as temporal reference points).

Due to the small amount of training data and the subjective natural of the dependent variable (cue strength score), we did not fully count on cross validation for the regression algorithms. We conducted another experiment to test the algorithms. In this experiment, a recognition and cued recall test were used to test the correlation of response time of recognition with the predicted score, and the likelihood of recognizing or recalling the event and the predicted score.

4.2.2 Findings

In the first experiment, we found that images tend to be much better cues than any textual information/attributes, especially digital photos, which far overweigh others in representing events, partly because manually taken digital photos are usually taken at the moments in which the lifelogger is more interested in, so that they want these moments to be captured actively. In fact, we found that the rating of importance of an event is highly correlated with the number of visits to photos from the event, the total number of photos, number of visits to the location of the event and the time spent in front of computers during the day. Following photos (including automatically captured photos), distinctive geo-location (e.g. city name), titles of computer activities are found to be good cues for general events.

In the second experiment (algorithm test), the cue strength score predicted by our algorithms was analyzed with the frequency of recognition and the response time of recognizing the represented event. Using an independent T-test, for most types of cues we found a significant advantage of predicted scores of cue strength for items that were recognized that those that were not recognized ($p<.05$). A comparatively strong correlation was also found between the response time and the predicted strength of image cues for both specific episodes and general events. Although the remainder of the correlations were not very strong, they were all positive.

In general, the positive results indicate that the algorithms are likely to select effective cues to represent events. They were integrated into our prototype system described in the next section, to select thumbnail images for event directories, landmark events, and facets in faceted browsing.

5. THE PROTOTYPE SYSTEM

The prototype interactive lifelog search system was designed based on the results of the two pre-development studies.

5.1 Front End Functions

The front-end of this system contained four main components: a search panel which serves the search function, a navigation panel which enables the user to navigate the folders for the target, a faceted browsing panel which allows them to filter and browse the results according to selected facets, and a result panel which displays the results. In this section, we introduce the system functions and explain the rationale behind the design.

Search panel: This provides rich selections of search fields, enabling users to search by content and/or context of the target items. This gives the user more freedom to search information with types of information that they remember. The following search fields are provided:

1) Content-based: Keywords, title, filename, item type, extension, and "from who/where" (path for files and URL for online items, "received from" and "send to" for emails and text messages).

2) Context-based:
When: The system provides flexible search options for "when" type queries. People seldom remember the exact numbers of date or time, they may remember or be able to infer the numbers for the approximate month (a week before Christmas), year, day of week (e.g. Fridays as they always have such meeting on Fridays), or part of the week (e.g. weekend). They may also be aware of approximate time based on their own schedules, e.g. after coffee, so that's around 3pm as the person always have coffee at around 3pm. For these reasons, we allowed plenty of flexibility to the query on date and time. In this system, we provided the following search options: year, season, month(s), part or day(s) of the week, part or hours of the day. Range sliders were available as an alternative to text format for people who remember the range.

Other Physical context: Where (Geo-Location of the individual: address, city, region, country), weather, light status and people appeared at the time of any instances of access or interacting with the target items.

The queries from the search panel are sent to an IR system in the backend, which uses a modified BM25F algorithm [23] to process string queries from the above mentioned fields.

Navigation panel: This adopts the folder metaphor, and automatically organizes data hierarchically by two attributes: time and location. It aims to cater for users who are used to hierarchical folder based navigation. Users can navigate and find items by: year→month→ week/date→ episode→files; or country →region →city →episode →files. Changing one attribute updates the directories in the other. For example, when navigating into the folder 2008→May, the location directories will only show directories of locations that the person visited during May 2008.

Faceted browsing panel: The content in this panel aims to mimic the structure of human autobiographical memory. It allows users to navigate and browse their collection as far as they can recognize the cues (attributes or images) of corresponding general events or episodes in which the target information has previously been encountered. Once a facet is selected (clicked), the directories in the navigation panel, as well as facets, change accordingly, to display the cues present in the general events or episodes which have the selected feature. The facet types include:

i) Images: images of important events in this collection.

ii) Location: representative country, region and city names.

iii) People: names of people that are representative of the events

iv) Contacts: name of email, SMS or IM contacts.

v) Computer activities: terms (titles and search terms) extracted from key computer activities.

Result panel: This displays search/faceted browsing results or a list of files in directories in the navigation panel. It also includes some result manipulation functions such as sorting and filtering. One of the episode memory supporting features in this panel is a timeline based browsing function. When the results are sorted by time, users can jump to adjacent areas of the results according to the date, time or landmark events on the timeline. Landmark events are represented by photos in the events, and were selected using an algorithm we developed in pre-development study 2 (section 4.2).

This system also allow users to search and browse "episodes", but these functions are beyond the scope of the discussion here.

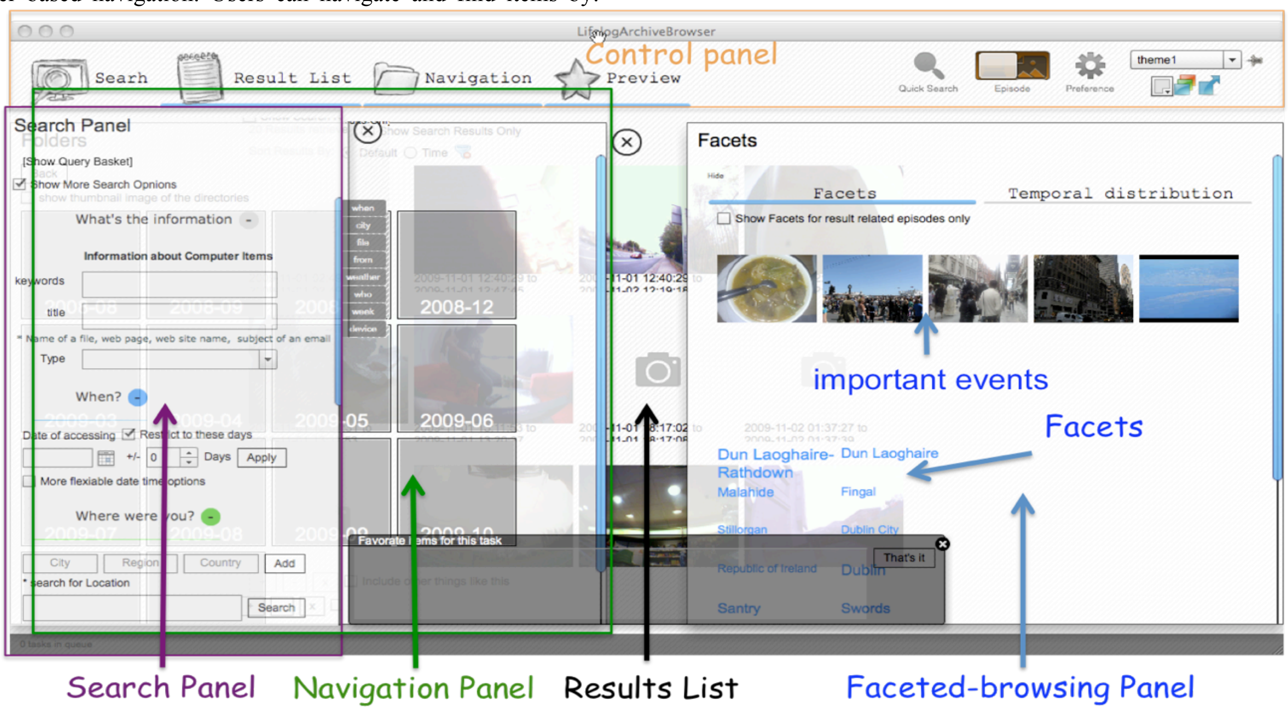

Figure 1. Components on the user interface of the prototype system.

5.2 Background Functions

The background functions were mainly realized through text processing technologies. To facilitate the process, the raw data were processed and stored in textual forms. The episodes were generated by segmenting data by timestamps according to context features, visual features of SenseCam images, and time intervals in computer activities [20]. Information of each episode was stored in an SQLite database with its timestamps, features (location, people appeared during the episode, weather) and keywords of computer activities during the episode. Each record of computer activity (including timestamps, title, attributes of the document or page, and full textual content, e.g.) was augmented with the attributes of its physical context, e.g. the location of user, name of nearby people, weather, Geo-location (country, region, city) and so on. Records of encountered items, episodes and attributes of images were stored in SQLite database. Computer items and episodes were also indexed to enable information retrieval function through an modified BM25F algorithm. Other functions such as navigation and faceted browsing were realized with in-house developed algorithms using data in the SQL database. For example, the terms (facets) and key images of landmark events were extracted based on the algorithm generated from the study described in Section 4.2.

6. EVALUATION OF EPISODIC CONTEXT FEATURES

After the prototype system was complete, we conducted a summative evaluation on the prototype system, with special focus on its effectiveness and efficiency for the refinding tasks.

6.1 Methods

The participants in this study were the three lifeloggers. Since these subjects had been unnaturally exposed to their data during pervious studies and during their own research, we left a one year gap between this finally evaluation and the last experiment, to let their memory, which would have reinforced by previous work and experiments, to fade to a comparatively natural level. This decision as based on a finding from one of our pilot exploratory studies [21]. We also noticed a dramatic improvement in the memory for the day in which the lifelog data was repeatedly exposed to the subjects, as the subjects remembered far more details of the day than any other days during the week. However, due to this one-year gap between the last date of lifelog collection and the starting date of the experiment, the content in their data collection is much less relevant to their current work, and therefore very unlikely to be needed in natural settings. For this reason, it will be difficult to collect enough usage data in natural setting. Besides, to test the effect of each feature, that is, to compare the performance of refinding tasks under baseline conditions and that under conditions when a certain feature was presented, we needed to be able to manipulate the tasks, in particular the level of difficulty for the tasks. Therefore, we adopted a "semi-natural approach". Instead of waiting for suitable tasks to arise, we required the subjects to actively generate a list of tasks for themselves.

6.1.1 Task Generation

In this study, our subjects generate the tasks in a comparatively natural way based on their own imagined scenarios of information needs. They were required to generate 4 tasks for each of the information types listed below:

1) Specific information: number, name of contact or papers, etc.;
2) Specific item: a file, an article, an email, an image;
3) Any information related to a topic: e.g. references on PIM;
4) Other types of information or items encountered on computers.

Sample scenarios were given as suggestions for generating tasks, e.g. *find some photos of you for use as profile images to upload to Facebook*, *find the name of the restaurant where the Christmas Lunch was held in 2009*. The participants were encouraged to generate as many tasks as they could before they were given the prototype system. They were required to add tasks using an in-house developed tool, which stored the task details in a database, so that they could be retrieved in a planned order during the evaluation experiments. To make the tasks as natural as possible, we wanted the subjects to make sure that their refinding tasks were carried for practical reasons. Therefore, we also required the subjects to tell us about:

1) Descriptions of the goal status: "A detailed description of the target without leaking any private information or information that you feel uncomfortable to reveal", "Describe in detail: What are you going to do with the target?"
2) Reason: describe how they were going to use the data, so that the last stage (use of results) could be evaluated.
3) Type of target: the choices are the 4 categories listed above. The selected type is used to evenly in assigning tasks for each condition.
4) Description of current memory status: this option is expected to indicate whether the subjects learnt more about their past while using the system, when compared with the search options in the system they actually used.
5) Planned approach for finding the target (before they have access to the lifelog system).
6) Expected difficulty of finding it with the approach described in the above question. This value is for comparison with the assessment in the post-task questionnaire, which asked the user about how difficult the finding task was when using the prototype system.

6.1.2 Task Assignment

To test each individual function, corresponding features were activated (enabled) to be tested one by one. The prototype system was modified to dynamically switch on and off a selected function, to allow the subject to complete their tasks under following conditions:

Baseline search interface (B1): The baseline search interface aimed to mimic a typical search interface, and therefore, it only included some typical search options in the search panel, e.g. keywords, titles, contact names, and the date range (e.g. 10 days around 05-06-2008). The result interface only included a list view together with basic sorting (Alphabet) and filters (item type).

Search interface with extended search options (C1): In this condition, contextual search options were added to the search panel, e.g. month, day of week, time of day, name of the location, people, weather and so on. The presentation of results remained the same as in the baseline system. There was an [unlock] button under the basic search panel, which could expand the interface to show the full list of search options. The subjects needed to click this button to use the extended search options. This small step aimed to avoid a ceiling effective of frequently using extended search options by discouraging subjects from using the them unless they really needed to. In this way, we could estimate how often subjects really needed to use the extended options.

Search interface with landmark assisted result browsing (C2): This condition enabled the function of using temporal context to assist the search result browsing. This condition was compared with the basic timeline-assisted browsing. In the baseline condition, the user could only jump to result items (which were previously encountered as recorded in their lifelog) according to calendar date on the timeline. we expected that the temporal landmarks could help users to recall or recognize the approximate time point to jump to browse and locate their target. Since the landmark events could act as good memory cues for autobiographical memory, the presentation of the landmarks may also trigger more memory of the past for refining search queries.

Baseline folder view (B2): This baseline interface included the labelled folder and a list view of results. The result panel, which presents all the items that belong to the folders and their subfolders. Users could either look for their targets in the result list with other assistive tools such as filtering by item type and sorting by time, or go further down the folder structure to see content in a more specific sub-collection (e.g. opening a month folder, or to open a week folder to view all events which happened in the week). This condition was compared to the baseline search interface. The advantage of folder-based navigation was expected to be in the consistency and people's habit of location-based storage and finding in the real world. Also, users could get rich contextual cues for recognizing and recalling the correct path.

Folder with contextual cues (C3): On top of the baseline folder view, richer contextual cues were included in this condition to assist with location-based navigation. These cues included cover photos (key images) of each folder, and the faceted browsing panel which displayed key images of events and a textual summary of the activities in the selected folder. The image and summary information were extracted using the algorithms which we developed from pre-development study 2 (Section 4.2). In this condition, the facets only serve an explanatory or cue function, and could not be clicked to update the collection as like in faceted browsing (C4). This condition was compared to condition B2. We anticipated that these cues could remind the subjects of the content in the folders, and more efficiently find the correct folder to navigate into.

Faceted browsing (C4): This condition included a faceted browsing panel and a list view of items. The facets were dynamically generated, using the algorithms we trained in pre-development study 2 (Section 4.2), to represent the currently selected collection, instead of the fixed hierarchy of the folder view (condition B2, C3), in which the user could only narrow down events by calendar units (month, week) or location (city, country).

Hybrid systems (H): This condition included all the components in the prototype system. The search panel, folder view and faceted browsing panel were presented by default. Other memory assistant functions, such as extended search options, landmarks, and contextual cues, were also available with a single click.

Table 1 below shows the features that are enabled in each of the above conditions, and the number of tasks that was to be assigned to each participant in each condition.

Table 1. Task and Enabled System features

Features	B1	C1	C2	B2	C3	C4	H
Basic feature options	Y	Y	Y				Y
Extended search options		Y					O
Basic result list	Y	Y	Y	Y	Y	Y	Y
Sort results by time	Y	Y	Y	Y	Y	Y	Y
Sort results by alphabet	Y	Y	Y	Y	Y	Y	Y
Alphabet landmark	Y	Y	Y	Y	Y	Y	Y
Temporal landmark events			Y			O	O
Folders				Y	Y		Y
Folder with key image					Y		O
Facets summary					Y	Y	O
Faceted browsing						Y	Y
Total tasks per subject	3	3	3	3	3	3	12

- *Y=enabled by default*
- *O=available, but the user needs to enable it manually, e.g. by click a button, or tick a checkbox.*

Since the features of the system were "unlocked" bit by bit, the subjects were introduced to the functions step by step. Therefore, only one or two components, elements or functions were introduced at a time. The subjects were allowed the opportunity to try each feature. Every time a new feature was unlocked (this happened before a task started), relevant instructions of how to use it appeared in a pop up window, with a link to a help file which introduced further details of the newly available functions. The subjects were encouraged to try the new features with an assigned sample task (assigned tasks were based on the examiner's knowledge of the items that should exist in all their individual personal lifelogs, e.g. received group emails). During the task, the subjects could use any available features in the experimental system to find the targets. They could select any amount of potentially usefully items into the results basket. On completing each task, they clicked a "That's it" button to indicate the completion of the task, and proceeded to the post-task questionnaire.

6.1.3 Evaluation Questionnaires

The system evaluation is based on a user activity log and pre- and post-task questionnaires. The purpose of the pre-task questionnaire was to present subjects with task details and enables them to finalize the target description, which would be visible to use. The post-task questionnaire was the main evaluation questionnaire, and contained the following questions:

1) Whether the target was found: subjects could select "yes", "partly", or "no".
2) The actual difficulty level of task, from 1=extremely easy to 5=very difficult. This rating was compared to the initial rating of expected difficulty of finding the target.
3) How they found the target or conducted the information-finding task if they did not find any relevant item. They were required to indicate the elements that they had used, and to rate how helpful each element was for this task. If they used the search function, they were asked to rate how difficult it was to formulate their query, and to leave comments if they had any regarding problems encountered in formulating it.
4) Reminiscing and emotional effects: did the information in the system remind them of any experiences in the past?
5) Reflection: did they learn anything about themself, or their life patterns while carrying out this task?

Question 4 and 5 were intended to explore the effect of accessing such a context rich personal information space.

6.2 Results

6.2.1 Episodic context as search options:

We logged the types of queries they used, and calculated the frequency of using each type of queries for 217 search sessions. A search session refers to a submission update to query. There were about 2.4 search sessions per task. We found that the most frequently used query type with the experimental interfaces was still keywords (50%). Apart from this, extension (29%), from contact (16%), date range (24%), day of week (27%), year (23%), month (27%), season (14%), country (18%), region (17%), city (15%) were also used comparatively often. Marginally used features include: path (4.6%), people nearby (3.7%), weather (3.6%), and light status (2.8%). In 30 of the tasks that the subjects could select to extend their default baseline interface to the advanced experiment interface, they only did so for 16 of them. This is partly because keyword-based search is usually more efficient with the background IR system. According to one subject, "I found the presentation easily through keyword search ... surprisingly it was one of the few results that appeared." While the extended options were not used as often as the basic search options, such as keywords or items type when looking for computer items, they are occasionally very helpful. Therefore, the low frequency of using the episodic contextual fields does not necessarily mean that they are not useful. In fact, A Chi-Square test found that when the extended search options were used, the subject was significantly more likely to find the planned target than when using the basic search options ($\chi2=4.43$, $p <.05$). We calculated the difference between the post tasks rating of difficulty and the rating of difficult rated when generating the tasks. A small but significant advantage was found ($p<.05$). This finding supports our first hypothesized method that *people are more likely to retrieve a target from an information retrieval (IR) system if they are allowed to generate a query with information from autobiographical context*.

6.2.2 Hierarchical Navigation

The system automatically generated two hierarchical structures of "folders" according to the time and location of events. Based on the findings of Bergman [22], we expected that people would tend to prefer this "location-based" finding method than search. However, in this study, our location and date-based directories were not favoured by our subjects. Subject A commented that she would rather that the "folders" were grouped by item type. This finding indicates that the mental "location" of computer items is still more closely associated with where one encountered the item previously, or how one usually finds it. For example, subject C commented for one of the tasks "I know where the file is on my computer, but I can't recall which month it was in", or in other words, the "mental categorization" of files is not usually based on calendar dates, but physical locations of the subject. However, this approach is very useful for finding photos, especially with the help of cover photos for each "folder", e.g. the subjects commented "folder navigation allowed me to find required photos, because I remembered the year and month", "folder navigation was great for this task because images from episode appeared in the folder". This difference in finding textual computer items and finding photos suggest that separate information management system or refinding functions might be designed for these two categories of personal information.

6.2.3 Faceted browsing

The faceted browsing function was intended as a supplement or an alternative to the hierarchical navigation function. It displayed the key facets and events in selected "folders". By clicking any of the facets, the collection was narrowed down to general events with the selected facet as a theme. Unfortunately, this function was seldom used alone, but usually as a filter function after search. Tasks in which faceted-browsing was used did not significantly improve the effectiveness of efficiency according to the post-task rating scores. Yet, the facets did successfully trigger more memories from the subjects, e.g. "the summary and photos brought that holiday back to mind".

6.2.4 Landmark Events on a Timeline

It is congruent to many other studies, e.g.[3], that our subjects also preferred to sort their search results by time. The timeline-based filter, which narrows results according to a selected time range, seemed to be a frequently used feature. Although we did not find an improved rate of finding the search target, nor did we find any significant decrease in difficulty level when this feature was enabled, the landmark events were commented to be useful. Among the 96 times that the timeline based filter was used (72 were under the task conditions that a landmark event is available), 57 of them were via a click on landmark events. This suggests that subjects do need the landmark events to more efficiently filter the results by time range.

6.3 Discussion

We evaluated three features in this system: 1) search with extended episodic context information, 2) navigation with rigid hierarchical folder structure according to calendar dates and name of locations, 3) browsing according to facets of corresponding episodes or general events. We combined both activity log and questionnaires to get the answers for the following questions:

1. *How often do users recall and use these features?*

As we can see from exploratory studies in Section 4.1, these features are quite often recalled when requested. However, during the tasks, the search fields for these "contextual attributes" were not used as often as they tend to be recalled (according to the records in previous study). This may not merely be an incidence of task difference or personal difference, but indeed, indicate that there is not a strong association between the likelihood of remembering an attribute and the likelihood of using this attribute to search. Since users are unlikely to exhaustively try all search options that they may have an answer, their choice of search options is largely influenced by factors such as: layout of the interface (using options on the top of the list), or experience of successful rate. In our case, this rate is negatively influenced by our imperfect data collection. For example, certain context information for some items was not captured, and therefore unable to be search by. Some computer activities were not recorded to temporally dysfunction of recording software. This makes the users fail to search to locate items by the temporal attributes of those activities. It is also possible that users are more likely to attend to the search options that they have in mind, that is, options that they are familiar with according to their past experience using other refinding and search systems. Unfortunately, we did not have enough tasks in this study to observe any trend of increasing use of search options relating to physical context.

2. *Does the episodes evoke users memory of textual content?*

This question was mainly assessed through the user's comments in the post task questionnaire. Although our users usually agreed and commented that the rich contextual information triggered vivid memories of relevant events, they never explicitly said that the information or images of the context (event) triggered their memory of textual information of the target.

3. *Autobiographical memory based file structure*

Both hierarchical navigation panel and faceted browsing functions were designed based on autobiographical memory. The user's comments suggested that calendar dates or geo-location (names of country or city) does not seem to be strongly associated with encounter information in the user's memory. This may also due to skipping the process of categorizing the items by automatically organizing objects. In fact, the process of organizing personal information is very meaningful as it integrates user's comprehension of the items. From the memory perspective, the association and the items is elaborated and enforced. In the future design, if we are to design personal information management system that provide folder (directory) based navigation function, perhaps we should think about how to support users to organize their items easily and in a way they can easily remember. For example, we may try to design a system that assists a user to organize their stuff into projects or events. Of course, the effectiveness of these methods is yet to be tested.

We expected that the facets which were extracted from computer activities could act as good cues to help users recognize the right "directory" for their target. Unfortunately, due to the limited number of tasks, our users did not have sufficient opportunities to use these functions. Therefore, we could not tell how effective the facets were, nor could we reject the hypothesis that the cues for events tend to trigger relevant memory about refinding targets.

7. CONCLUSIONS

In this paper, we reported a proof-of-concept case study, which investigated the effectiveness of embedding features that utilize and support users' episodic memory during a refinding task. The features were built on models from cognitive psychology literature about human memory, which suggests that physical context is likely to be associated with people's memory of encountered information, and may trigger more memory of the information, or be triggered and recalled by information about the target. We proposed three approaches that make use of autobiographical memory and memory of the physical context: i) provide search fields which allows users to search by features from the physical context, ii) dynamically organize personal information space in a structure similar to autobiographical memory, iii) provide users with cues for potentially relevant general events. Following the guidelines, we carried out a series of studies, which collected concrete attributes and quantities for developing a working system which provided the proposed functions. In these studies, we also found evidence that physical context is likely to be recollected for refinding tasks.

The prototype system, which we developed based on the above investigation, provided four functions that support or utilize autobiographical memory and memory of physical context: extended search function with options about physical context, date and location based hieratical folder structure for navigation, a faceted browsing function based on autobiographical memory structure and provides cues for events, and a landmark augmented timeline for browsing. Among these features, the extended search option was the most frequently used, although our subjects tended to use the keyword based search even more often. The rigid date or location hierarchical directory structure was not very helpful for finding textual items, due to the loose connection of exact date or geo-location and accessed computer items. Although the faceted browsing and timeline functions were not used frequently, they were commented to be evocative. Based on the statistics and the subjects' comments, the memory-supporting function in this prototype system did reduce the difficulty of finding tasks and improve the likelihood of finding the information in personal information space. This is congruent with our main hypotheses on the utility of context information in refinding tasks. The findings of this study also suggest that separate systems or features would be better for life related perceptual rich items such as photos, and for textual rich items as in traditional personal information spaces.

In this case study, we only managed to hire three participants in the final study, due to both physical and mental burden of wearing necessary devices to collect contextual data from context. The small sample size makes it difficult to provide generalizable results. With the development of sensor and smart phone technologies, it will become much easier to collect such data seamlessly. We also suffered from many technical problems which reduced the usability of the system. This caused a number of refinding tasks to fail. In the near future, we can construct more robust systems with many more potential subjects to explore these questions, to achieve more solid results.

8. ACKNOWLEDGEMENT

This work was supported by Science Foundation Ireland Research Frontiers Programme 2006. Grant No: 06/RFP/CMS023, and Grant No: 07/CE/I1142 as part of the Centre for Next Generation Localisation (CNGL) project at Dublin City University.

9. REFERENCES

[1] Jones, W. P., and Teevan, J. 2007. *Personal information management* .University of Washington Press.

[2] Lansdale, M. W. 1988. The psychology of personal information management. *Applied Ergonomics, 19*(1), 55-66.

[3] Cutrell, E., Dumais, S. T., & Teevan, J. 2006. Searching to eliminate personal information management. *Communications of the ACM, 49*(1), 58-64. .

[4] Elsweiler, D., Baillie, M., & Ruthven, I. .2008. Exploring memory in email refinding. *ACM Transactions on Information Systems (TOIS), 26*(4), 1-36.

[5] Blanc-Brude, T., & Scapin, D. L. 2007. What do people recall about their documents?: Implications for desktop search tools. Paper presented at the *Proceedings of the 12th International Conference on Intelligent User Interfaces,* Honolulu, HI, USA. 102-111.

[6] Elsweiler, D., & Ruthven, I. 2007. Towards task-based personal information management evaluations. In *Proceedings of the 30th ACM SIGIR*. ACM. pp 23-30.

[7] Gonçalves, D., & Jorge, J. 2005. "Tell me a story" issues on the design of document retrieval systems. *Engineering Human Computer Interaction and Interactive Systems*, 117-134.

[8] Kelly, Liadh, et al. "A study of remembered context for information access from personal digital archives." *Proceedings of the second international symposium on Information interaction in context*. ACM, 2008.

[9] Ringel, M., et al. 2003.Milestones in time: The value of landmarks in retrieving information from personal stores. In *Proceedings of Interact 2003, the Ninth IFIP TC13 International Conference on HCI*. 2003. Zurich, Switzerland: IOS Press Inc.

[10] Tulving, E., & Thomson, D. M. 1973. Encoding specificity and retrieval processes in episodic memory. *Psychological review*, *80*(5), 352.

[11] Anderson, J. R., & Bower, G. H. 2013. *Human associative memory*. Psychology press.

[12] Anderson, J. R., Bothell, D., Byrne, M. D., Douglass, S., Lebiere, C., & Qin, Y. 2004. An integrated theory of the mind. *Psychological review*, *111*(4), 1036.

[13] Smith, S. & Vela, E. 2001. Environmental context-dependent memory: A review and meta-analysis. *Psychon. Bull. Rev.* **8** (2): 203–220.

[14] Baddeley, A. D., Eysenck, M. W., & Anderson, M. 2009. *Memory* Psychology Press.

[15] Conway, M. A., & Pleydell-Pearce, C. W. 2000. The construction of autobiographical memories in the self-memory system. *Psychological Review, 107*(2), 261.

[16] Friedman, W. J. (1993). Memory for the time of past events. Psychological bulletin, 113(1), 44.

[17] Horvitz, E., Dumais, S., & Koch, P. 2004. Learning predictive models of memory landmarks. In *Proceedings of the CogSci 2004*. HCI International.

[18] Elsweiler, D., & Ruthven, I. 2007. Towards task-based personal information management evaluations. In *Proceedings of the 30th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval,* 23-30.

[19] Chen, Y., & Jones, G. JF. 2013. Utilizing Episodic Context in Search of Personal Information Archives. In *proceeding of Irish HCI 2013, Dundalk, Ireland*.

[20] Chen, Yi. 2013. *Communicating with your E-memory: finding and refinding in personal lifelogs.* PhD thesis, Dublin City University.

[21] Chen, Y., & Jones, G. J. F. (2009). Exploring digital elements for visualizing time in personal information refinding. *In proceedings of 3rd Irish HCI Conference*, September 2009, Dublin, Ireland

[22] Bergman, O., et al. 2008, Improved search engines and navigation preference in personal information management. *ACM Trans. Inf. Syst.*, 2008. 26(4): p. 1-24

[23] Kelly, Liadh (2011) *Context driven retrieval algorithms for semi-structured personal lifelogs.* PhD thesis, Dublin City Universiy