The Role Of Places And Spaces In Lifelog Retrieval

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

Daragh Byrne
CLARITY: Centre for Sensor Web Technologies
Dublin City University
Dublin 9, Ireland
dbyrne@computing.dcu.ie

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

ABSTRACT

Finding relevant interesting items when searching or browsing within a large multi-modal personal lifelog archive is a significant challenge. The use of contextual cues to filter the collection and aid in the determination of relevant content is often suggested as means to address such challenges. This work presents an exploration of the various locations, garnered through context logging, several participants engaged in during personal information access over a 15 month period. We investigate the implications of the varying data accessed across multiple locations for context-based retrieval from such collections. Our analysis highlights that a large number of spaces and places may be used for information access, but high volume of content is accessed in few.

1. INTRODUCTION

Personal lifelog archives [8] can contain data from many diverse sources, e.g. personal photos, mobile phone SMSs, emails, IM chats, documents created, web pages viewed, etc, created within diverse information access spaces, including home, work, and social locations. Such collections are automatically and passively collected as a user goes about their day-to-day activities thereby offering rich insights into their lives. Given the volume and diversity of content within these archives, there is a clear challenge in the retrieval and location of important and relevant items in response to user queries, and also in presenting interesting data to a subject browsing through their archive. Any additional information which can assist in identifying important items is thus potentially very important. One potential source of useful information is contextual metadata which can be associated with the individual items within a long-term lifelog. One such contextual cue which can be applied is the location of the individual - the spaces in which they use this information. In this paper we explore the role of these spaces within information access as a cue to the potential utility of automatically captured geo-location context data within long-term lifelogs.

Personal Information Management: PIM 2009 ASIS&T 2009 Workshop, November 6-7, 2009, Vancouver, British Columbia

2. BACKGROUND & MOTIVATION

Lifelogs may contain many sources of automatically captured digital content - everything from items read, written, or downloaded; to footage from life experiences, e.g. photographs taken, videos seen, music heard, details of places visited, details of people met, etc, along with details of passively captured location and social context. This context data is important within lifelogs and as such shows its potential utility in retrieval scenarios [6], [14]. The notion of using context to aid retrieval in this and other domains is not new [12], [11] and existing work, such as [2], [5] and [17], have used context data such as location of file, actions performed on file, daylight status, weather and local time to aid file retrieval. Explorations which use manual annotation, in the form of *tagging*, have also been conducted to provide contextual metadata such as people present [4]. Tagging however places undue burden on the user, it is more desireable to employ wholly automatic techniques to provide such information. This has been outlined in our previous work [13].

Current studies in personal information retrieval tend to focus on one type of data, e.g. file, image or mobile data, retrieval. Rich personal lifelogs however are neither confined to the desktop, mobile, audio or image spaces, and are likely to contain data pertaining to the various facets of an individual's life, e.g. work and personal data. Our work emphasises the capture of multiple complementary sources in tandem and is discussed in the subsequent sections. We postulate that the types of context data which will prove beneficial to retrieval may vary across the many personal spaces contained in these rich multi-modal lifelogs. In particular in this paper we explore the utility of recalled geo-location in narrowing a lifelog search space.

3. EXPLORATION

The work presented here is an initial exploration of a single contextual channel, geo-location of the collection owner, and the potential value it adds to the lifelog. While location may be well recalled [14], it is our expectation that the value of context, and in particular location, within retrieval will vary across spaces. In this initial investigation we analyse the relationship between content in the lifelog and the location in which it was created, accessed or reviewed.

3.1 Collection Overview

For the past 15 months we have been engaged in continuous large scale multi-modal lifelog build-up for three subjects. These subjects (1 male, 2 females) are all post-

graduate students within our University. Subject 1 lives in a different region/county to the University. Subjects 2 and 3 live in close proximity to the University. All participants travel to some extent for leisure and research purposes. Our current archives of personal data annotated with rich sources of automatically generated context are much larger and more heterogeneous than typically created digital archives. These lifelogs contain data from a variety of complementary sources, including:

- Desktop Activity: All laptop and PC activity is monitored (every item (email, word document, web page, etc) accessed by the user, with time and duration of access, contents of item, path to item information, etc) for subjects using a combination of MyLifeBits [7], Slife [16] and in-house scripts.
- Passive Capture Media: Continuous passive image capture is enabled through the use of a small wearable device, the Microsoft Research SenseCam [9].
- Mobile Activity Data: Mobile phone activity in the form of call logs are recorded using a proprietary piece of software and SMSs are captured using an application developed in-house.
- Mobile Context Data: This data is captured on subjects' Nokia N95 mobile phones by constantly running the Campaignr software, provided to us by UCLA (USA) [10]. This provides location cues of GPS data, wireless network presence and GSM location data, from which placenames, light status and weather conditions can be derived, and co-present Bluetooth devices, from which people present can be uncovered [15], [1].
- Biometric Information: Physiological and heart rate readings taken from wearable biometric devices [3], which allow us monitor physiological conditions and infer emotional state were continuously captured for a one month period.

For this current preliminary investigation the exploration presented is restricted to PC, laptop and mobile phone activity - in particular SMS's sent and received[1], web pages viewed, emails created or accessed, and computer files created or accessed on laptop and PC for the 15 month lifelog capture period. While we acknowledge that people may consume other forms of textual data on mobile phones, SMSs were the only textual digital content consumed by our subjects. Additionally across the entire populous, other types of digital devices are used by individuals, for example PDAs. Our test subjects did not use such devices.

It should also be acknowledged that for geo-location logging due to device crashing[2], subjects' need to conserve battery life, subjects forgetting to turn on the device and subjects' occasional need for privacy some periods of the 15 month textual lifelogs are not annotated with geo-location. For Subject 1, 73% of their activity is geo-location tagged, for Subject 2 34% is geo-location tagged, and for Subject 3 83% is geo-location tagged.

[1]It was not possible to capture individual accesses to SMS messages using currently available software, hence only time of SMS sending or receiving is captured.

[2]During the first 8 months of the logging, the infrastructure and software was still naesent undergoing iterative change and improvement, in particular the mobile logging software. During this time device & software crashes occured intermittently and went undetected until the subject checked if the software was still running. Following the first 8 months a much more stable platform was available resulting in very occasional device crash.

Table 1: Total number of item creation/accesses captured by subjects on PC, laptop and mobile phone.

Item Type	Subject 1	Subject 2	Subject 3
Laptop			
Webpages	7105	3096	19894
Email	247	542	209
Word Document	2412	0	30
Excel File	2182	22	77
Powerpoint	886	0	18
PDF	663	155	98
Text File	692	1	184
XML File	0	166	18
Code / Language	4768	4	1562
Media	229	16	16
Other	13676	1359	3160
Laptop Total	*32860*	*5361*	*25266*
PC			
Webpages	3494	15781	78375
Email	554	618	11647
Word Document	992	2238	4885
Excel File	1122	406	1529
Powerpoint	537	558	677
PDF	454	1089	394
Text File	401	146	1407
XML File	82	180	685
Code / Language	4421	645	18450
Media	104	13	502
Other	16736	17149	26442
PC Total	*28897*	*38823*	*144993*
Mobile Phone			
SMS	3232	436	3023
Phone Total	*3232*	*436*	*3023*
Total	**64989**	**44620**	**173282**

3.2 Discussion - Collection Contents

As mentioned above, this evaluation is based upon the PC, laptop and SMS content collected within the lifelogs of three individuals over a fifteen month period. Over the course of this period, an individual will encounter a broad range of unique digital content which may be viewed or accessed in isolation or reviewed periodically. To better understand both the content of a lifelog and the prevalence of the various content types the users worked with, Table 1 provides a breakdown of information access across the available devices[3]. It illustrates that there is a huge dominance within the digital landscape of a lifelog for information creation/access to webpages, communication via email[4] (for Subject 3) and SMS messages (for Subjects 1 and 3) and code development (for Subjects 1 and 3).

Taking the information presented in Table 1, serves to

[3]Data type 'other' in Table 1 represents such things as file system accesses for example.

[4]Note for Subjects 1 and 2 little information on email accesses was captured owing to limitation in software.

highlight the different relationships the individuals have with the content housed within their lifelogs. We can see that these individuals have very different personal information management and access strategies. For example, participant 3 consumes high volumes of web pages and is more engaged in coding/development as compared with the other participants. In fact 69% of this participants computer activity is spent engaging in these tasks. While participants 1 and 2 consume more than double the volume of pdf documents of participant 3. Further to this we can also see that the devices employed by the participants have very different roles - with participant 1 favouring their laptop while participants 2 and 3 display a strong affinity to their desktop computer for content access. In particular, participant 1 greatly favours their laptop for Web page access and Word and Excel file creation or access.

3.3 Discussion - Geo-location Tagging

It is particularly important to consider the mobility of these devices as this is extremely pertinent to the utility of location-based context in narrowing the search space in such collections. For example, we can expect a desktop computer to be almost stationary in its location for all of its use (perhaps occasionally it might be moved but this is likely to be a rare event), while a laptop is mobile and therefore can be expected to traverse several locations in its use. In this data analysis we set out to examine the extent to which recalled geo-location can narrow a lifelog search space using the PC, laptop and mobile data in our subjects' lifelogs. The life styles of our subjects afford them certain levels of movement between geo-locations as described in Section 3.1. However, it is acknowledged that while the patterns of movement of our subjects are typical of those of many individuals, they do not represent the entire populous. This study seeks to form an initial exploration of the utility of geo-location in narrowing down the search space for individuals with relatively stationary lifestyles.

The nature of the device is likely to have implications for its relationship to the spaces in which it operates, this is illustrated within Tables 2 and 3. The tables present a list of the volumes of laptop, PC and mobile activity encountered in each 'place' that the subjects were determined to be in over the 15-month period for geo-location tagged items within each subjects lifelog. Here we consider a 'place' to be a unique region + county + city within a given country that the individual was located in. These 'places' were extracted from the available GPS data. Locations have been anonomysed, however encounters within the same country can be discerned by the use of the same prefixing letter.

We explored 'places' of the granularity of country + region + county + city + street. However, when we considered location at the granularity of street level several false locations were noted. Using GPS location alone poses issues with accuracy and achieving sufficient granularity to be useful. However, in the future using detected wifi and bluetooth signals may help overcome this.

As assumed a desktop will largely be confined to a single location as can be seen in Tables 2 and 3. For Subjects 2 and 3 the laptop is also used predominately in just a single location. In contrast to Subjects 2 and 3, Subject 1 has multiple dominant locations for laptop activity. For all subjects it does pentrate many more 'places' but to a lesser extent as shown in Tables 2 and 3. As might be expected, the mobile SMS content is most widely distributed across spaces for all subjects.

Given this information, what implications does it bear for retrieval using context? First the desktop should be considered. Large volumes of information are accessed in principally one static location. This suggests that in situations where an individual seeks to identify an item of interest from their PC that (location-based) context information may offer little assistance. Attempting to weight or filter by location in this circumstance is unlikely to sufficiently narrow the search space or clearly identify an item of relevance. The same will similarly be true for laptop activity for the subjects who have a dominant location for laptop activity. However, in situations where an individual does not recall the device the item was created or accessed on, recalling the dominant geo-location will narrow the search space, albeit not to the same extent as recollection of an item created or accessed in a non-dominant location. Conversely, given that some laptop accesses occur at infrequently encountered locations, when retrieving for such items context may offer real utility. The lack of one dominant location for laptop activity observed for Subject 1 suggests that geo-location will be of greater utility in narrowing the search space for this subject. Subjects 1 use of different spaces for laptop activity also serves to highlight the fact that variations in behaviour can be expected for different subjects.

Finally the spread of encountered geo-locations for sent and received SMSs suggests that location recall would be effective in aiding identification of required SMSs for all three subjects. This finding also suggests that geo-location offers general utility for any mobile phone content, for example photos, tweets, etc.

These observations suggest utility for the inclusion of geo-location context retrieval facilities in lifelogging retrieval systems. However, investigation of the volumes of geo-location recalled for retrieval scenarios and experiments to determine its 'real' utility in improving the detection of required items in lifelog search systems is required. The analysis carried out in this paper provides initial support for investing in such analysis.

4. CONCLUSIONS AND FUTURE WORK

This paper provides an initial exploratory investigation of the role of context data within a lifelog collection and its potential utility in the retrieval of content from these collections. The variation in types of content within a lifelog was illustrated. We then examined the relationship between this content and the 'places' in which it was accessed, through location-context gathered as part of the collection. The relationship between the space of access and the mode of access was further probed. This served to highlight the often implicit relationship between space, place, access medium and personal information. It is clear from the results presented that information access is often confined to one or two dominant spaces, however, mobile access - via laptops or phones - increases the spaces in which our personal media can penetrate. Space or place of access thus has implications on how we might later seek to retrieve that content, particularly should we employ context-cued or -aware approaches.

While we cannot generalise about the laptop, PC and mobile habits of entire populous from the observations made in this paper, the results give insight into the long term laptop, PC and mobile activity for individuals with relatively sta-

tionery lifestyles. They also highlight, even with our small subject set, that variation in the use of spaces is to be expected across individuals.

As part of our future work, we will undertake more detailed exploration to examine how the role of space can play a part within lifelog content retrieval and indeed how our relationship to that space should inform the retrieval place. We additionally plan to further explore how tacit and accurate location cues are to the individual in relation to the dominance of that space within the landscape of their lifelog content.

5. ACKNOWLEDGEMENTS

This work is funded by grants under the Science Foundation Ireland Research Frontiers Programme 2006 and under grant 07/CE/I1147. We would also like to thank the Irish Research Council for Science, Engineering and Technology.

6. REFERENCES

[1] B. Lavelle and D. Byrne and C. Gurrin and A.F. Smeaton and G.J.F. Jones. Bluetooth Familiarity: Methods of Calculation,Applications and Limitations. In *MIRW 2007 - Mobile Interaction with the Real World, Workshop at the MobileHCI07: 9th International Conference on Human Computer Interaction with Mobile Devices and Services*, 2007.

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

[3] BodyMedia - http://www.bodymedia.com.

[4] E. Cutrell, D. Robbins, S. 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.

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

[6] M. Fuller, L. Kelly, and G. Jones. Applying contextual memory cues for retrieval from personal information archives. In *PIM 2008: Proceedings of Personal Information Management, Workshop at CHI 2008*, April 2008.

[7] J. Gemmell, G. Bell, R. Lueder, S. Drucker, and C. Wong. MyLifeBits: Fulfilling the Memex Vision. In *ACM Multimedia '02*. Juan Les Pins, France, December 2002.

[8] J. Gemmell and H. Sundaram, editors. *CARPE 2004 The First ACM Workshop on Continuous Archival and Retrieval of Personal Experiences*, 2004.

[9] 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 2006*, pages 177–193, 2006.

[10] A. Joki, J. Burke, and D. Estrin. Campaignr - a framework for participatory data collection on mobile phones. Technical report, Centre for Embedded Network Sensing, University of California, Los Angeles, October 2007.

[11] G. Jones. Challenges and Opportunities of Context-Aware Information Access. In *Toshiba Corporation Research and Development Center*. Kawasaki, Japan, 2005.

[12] G. Jones and P. Brown. The Role of Context in Information Retrieval. In *SIGIR 2004 - the 27th Annual International ACM SIGIR Conference, Workshop on Information Retrieval in Context*, 2004.

[13] L. Kelly. The information retrieval challenge of human digital memories. In *BCS IRSG Symposium: Future Directions in Information Access 2007*, 2007.

[14] L. Kelly, Y. Chen, M. Fuller, and G. 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.

[15] V. Kostakos, E. O'Neill, and S. Jones. Social networking 2.0. In *Extended abstracts of the conference on Human factors in computing systems*, pages 3555–3560, 2008.

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

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

Table 2: Total SMS, laptop and computer activity across geo-locations for subjects.

Location				Subject 1			Subject 2			Subject 3		
Country	**Region**	**County**	**City**	**Laptop**	**Mobile**	**PC**	**Laptop**	**Mobile**	**PC**	**Laptop**	**Mobile**	**PC**
A	1	a	1							84		
A	2	a	1								1	
B	1	a	1		1							
B	2	a	1		1							
C	1	a	1								10	
C	1	a	2								6	
C	1	a	3								100	
C	2	a	1								3	
C	2	a	2								25	
C	2	a	3								2	
C	2	a	4								31	
D	1	a	1		23							
E	1	a	1					5				
F	1	a	1					1				
F	2	a	1					2				
F	2	b	1	2								
F	2	b	2	1								
F	2	b	3	6								
F	2	c	1		1			1				
F	2	d	1	258	13			4				
F	3	a	1					1				
F	4	a	1					1				
F	5	a	1								18	
F	6	a	1					1				
G	1	a	1	3	2						10	
G	2	a	1		1							
G	3	b	1		7							
G	3	b	2		1							
G	4	a	1	358	314	16918	337	15	12391	10212	1407	117712
G	5	a	1		3						2	18
G	6	a	1	11	3	48					4	
G	6	a	2	25	9	94						
G	6	b	1	4	18	82						
G	6	b	2	1	2	15						
G	6	b	3	18	13	136					11	
G	6	c	1					1				
G	7	a	1	10								
G	7	a	2	2								
G	7	a	3	2								
G	7	a	4	2								
G	7	a	5	42	5							
G	8	a	1	8	3							
G	8	a	2	9	2							
G	9	a	1	2								
G	9	a	2	2								
G	9	b	1				1					
G	10	a	1							46	7	
G	10	b	1								4	
G	11	a	1								2	
G	11	b	1								1	
G	11	c	1	3								
G	11	c	2	10	1							
G	11	c	3		8						10	
G	11	c	4	9065	549	2339				912	239	
G	11	c	5		1							
G	11	c	6	6460	451	323					1	
G	12	a	1		1							
G	12	b	1	1336	66	2668	71	14	2119	1853	49	5950
G	12	c	1							1570	85	3011
G	12	d	1							6		

Table 3: Continuation of Table 2 - Total SMS, laptop and computer activity across geo-locations for subjects.

Location				Subject 1			Subject 2			Subject 3		
Country	**Region**	**County**	**City**	**Laptop**	**Mobile**	**PC**	**Laptop**	**Mobile**	**PC**	**Laptop**	**Mobile**	**PC**
G	12	e	1		1		1					
G	12	e	2				56					
G	12	e	3				11	3				
G	12	f	1		1							
G	12	f	2		2		2				2	
G	12	g	1		1							
G	12	h	1		1							
G	12	i	1	72								
G	12	j	1	238	11						10	
G	12	j	2	4765	237							
G	12	k	1				1					
G	12	l	1				2					
G	13	a	1				10					
G	14	a	2		3		5	3				
G	15	a	3		1							
G	15	a	4	1							2	
G	15	a	5								7	
G	16	a	1	2								
G	16	b	1	4	4							
G	17	a	1	2								
G	18	a	1				4					
H	1	a	1	35	18	8	1	3	103		8	36
I	1	a	1		15							
I	1	b	1		1							
J	1	a	1							718	8	
J	1	a	2							2	84	
J	2	b	1								1	
			Total	*22759*	*1795*	*22631*	*502*	*55*	*14613*	*15403*	*2150*	*126727*