Hideo Joho
Frank Hopfgartner
Cathal Gurrin(Eds.)

Information Retrieval and Learning with Lifelogging Devices

– A Session for Interaction and Engagement at iConference 2016 –

March 18, 2016

DCU
Volume Editors

Cathal Gurrin
Dublin City University
E-mail: cgurrin@computing.dcu.ie

Hideo Joho
University of Tsukuba
E-mail: hideo@slis.tsukuba.ac.jp

Frank.Hopfgartner
University of Glasgow
Frank.Hopfgartner@glasgow.ac.uk

Preface

We are pleased to present the proceedings of Information Retrieval and Learning with Lifelogging Devices - A Session for Interaction and Engagement. The session took place on 21 March 2016 as full day event of iConference 2016 in Philadelphia, PA, USA. The iConference conference series is established by the iSchools organisation, a worldwide association of Information Schools.

The main focus of the session was to identify opportunities of lifelogging devices such as wearable devices and mobile applications for researchers, developers, learners, and educators in the field of information science. Such devices and apps have become commodity items, and collecting a massive amount of personal lifelog data has become feasible for everyone, thus offering new opportunities and challenges to advance research in the domain. For example, information seeking research has traditionally relied on questionnaire or interview data to understand the contextual factor in depth, yet showed limited quantitative verification of their theories. Wearable devices can provide a strong evidence to support mixed-design studies. Learning analytics have also been a popular subject, and many applications have been developed using MOOC data. However, we do not have much insights from students lifelogging activities. Therefore, there are still many gaps to bridge among researchers, developers, learners, and educators, to fully leverage the power of lifelogging devices and their data.

Addressing these gaps, we aimed for an interactive session where participants could show demo systems and discuss early research outcomes or novel ideas about how to use lifelog data. Participants were asked to submit research papers that describe their systems, study showcases, or interesting ideas about lifelogging devices or lifelog data. The submitted papers were reviewed by the organisers and PC members.

The session was separated into three parts. In the morning session, the organisers shared their experiences of lifelogging over almost a decade. They presented various lifelogging devices, including cameras, activity trackers, and desktop/mobile apps. The aim of this session was to create a living lab environment where participants could experience these lifelogging. In the first session in the afternoon, three oral presentations about lifelogging technologies and evaluation were given. We first saw a presentation of Grefenstette on The Place of Text Data in Lifelogs, and Text Analysis via Semantic Facets. Gurrin then presented A ColourMap Visualisation of Lifelog Data, followed by giving an introduction to the NTCIR12-Lifelog: A Test Collection to Support Collaborative Benchmarking of Lifelog Retrieval. In the final session, Joho presented two case studies on the use of lifelogging: The first presentation was about Lessons Learned from 8 Months of
Lifelogging: Case of 2nd Year Undergraduate Students at the University of Tsukuba, Japan. In the second presentation, he talked about a case study on Lifelogging by Senior Citizens in a Highly Ageing Society. The session ended with a round table discussion to identify some of the core research directions regarding the development and use of lifelog devices in Information Research and Learning.

We would like to thank the many people who contributed to the success of this session. We deeply appreciate the hard efforts of the organisers of iConference who were very flexible and helpful during the preparation of this session. Moreover, we would like to thank our programme committee who agreed to review the submitted papers. This includes: Graham Healy (Dublin City University, Ireland), Havard Johansen (UIT - The Arctic University of Norway, Norway), Na Li (Dublin City University, Ireland), Mamiko Matsubayashi (University of Tsukuba, Japan), Atsushi Matsumura (University of Tsukuba, Japan), Michael O’Mahony (University College Dublin, Ireland), Amon Rapp (University of Torino, Italy), Tetsuji Satoh (University of Tsukuba, Japan), Yohei Seki (University of Tsukuba, Japan), Masao Takaku (University of Tsukuba, Japan), Taro Tezuka (University of Tsukuba, Japan), Robert Villa (University of Edinburgh, UK), and Jiang Zhou (Dublin City University, Ireland). Last, but not least, we would like to thank all authors who submitted their work to this session, and everyone who contributed to the session in any way.

Dublin, Ireland, March 2016

Hideo Joho
Frank Hopfgartner
Cathal Gurrin
Tables of Content

On the Place of Text Data in Lifelogs, and Text Analysis via Semantic Facets. .............................................................................................................. 1
Gregory Grefenstette, Lawrence Muchemi INRIA, Saclay

A ColourMap Visualisation of Lifelog. .......................................................... 4
Aaron Duane, Moohamad Hinbarji, Rami Albatal, Cathal Gurrin

NTCIR12-Lifelog: A Test Collection to Support Collaborative Benchmarking of Lifelog Retrieval Systems. .......................................................... 7
Liting Zhou, Cathal Gurrin, Rami Albatal, Hideo Joho, Frank Hopfgartner

Lifelogging by Senior Citizens in a Highly Ageing Society: A Pilot Study. . . . 9
H. Joho, M. Matsubara, N. Uda, K. Donkai, and C. Mizoue

Lessons Learned from 8 Months of Lifelogging. ............................... 13
Natsuki Nagaoka, Yuzuki Furuhashi, Naoko Minei, and Hideo Joho
On the Place of Text Data in Lifelogs, and Text Analysis via Semantic Facets

Gregory Grefenstette\textsuperscript{1}, Lawrence Muchemi\textsuperscript{2}
\textsuperscript{1}INRIA, Saclay, France
\textsuperscript{2}University of Nairobi, Kenya & INRIA, Saclay, France

Abstract

Current research in lifelog data has not paid enough attention to analysis of cognitive activities in comparison to physical activities. We argue that as we look into the future, wearable devices are going to be cheaper and more prevalent and textual data will play a more significant role. Data captured by lifelogging devices will increasingly include speech and text, potentially useful in analysis of intellectual activities. Analyzing what a person hears, reads, and sees, we should be able to measure the extent of cognitive activity devoted to a certain topic or subject by a learner. Test-based lifelog records can benefit from semantic analysis tools developed for natural language processing. We show how semantic analysis of such text data can be achieved through the use of taxonomic subject facets and how these facets might be useful in quantifying cognitive activity devoted to various topics in a person’s day. We are currently developing a method to automatically create taxonomic topic vocabularies that can be applied to this detection of intellectual activity.

Keywords: Lifelog; Taxonomy Induction; Semantic Analysis; Cognitive activities

Citation:
Gregory Grefenstette, Lawrence Muchemi.
Contact: gregory.grefenstette@inria.fr, lawrence.githiari@inria.fr.

1 Introduction

The overarching goal of lifelogging is the creation, storage, analysis and eventual usage of digital records of an individual’s total experience. In practice, it currently involves the passive digital capture of moments and episodes in an individual’s everyday life. One objective of lifelogging is increasing self-awareness and eventual improvement of one’s life. The quantity of wearable lifelogging gadgets, which track activity, physiological and environmental data, continues to grow and expand. Alongside these quantified-self devices, body-worn video devices, e.g. helmet cameras, smartglasses, which capture image and sound, are beginning to appear. In the near future, the images they capture can be transformed into text via optical character recognition via object recognition software. Captured sound will be converted to text via automatic speech recognition. Already, people process great quantities of text every day. These emails, social network posts, and web pages visited can also be passively captured and processed for an individual’s lifelog (Hinbarji, et al., 2016).

Current value-added efforts for lifelog data concentrate on analysing physical activities as opposed to cognitive activities. Analysis leading to better understanding of an individual’s daily information context would benefit many research fields such as self-directed learning in online e-learning platforms. For example, a conclusion arrived at in self-directed learning environments by Guralnick (2007) establishes that an individual’s information context “influences the level of learner autonomy that is allowed in the specific context, as well as how a learner utilizes resources and strategies, and becomes motivated to learn”.

This position paper seeks to explain how abstract text-based environments, capturable in lifelogs, might be analysed through taxonomic facets that characterise and quantify the areas of intellectual activities that a person engages in their daily life.

2 Present and Future of Textual Lifelog Data

In addition to quantified-self data (heart rate, steps taken, liquids and food consumed, mood, arousal, blood oxygen levels, sleep), a lifelog can increasingly contain text, sounds and images. The text to be included in a lifelog can come from four main sources: (i) digital interactions such as emails sent and received, social network posts, documents stored on a computer, web pages visited; (ii) conversion of captured, ambient speech into text via automatic speech recognition; and (iii) conversion of printed text via optical character recognition (Yi & Yingli, 2015); and (iv) the conversion of GPS coordinates into semantic descriptions of places visited (Xin, Cong & Jensen, 2010). Some numbers: A business user will receive about 75 legitimate emails per day, and send over 30 (Radicati & Levenstein, 2015). The average
online user consumes over 280 posts per day amounting to 54,000 words (Bennett, 2013; Dhir & Midha, 2014). As passive conversion of speech to text continues to improve, the quantity of text to be stored on a lifelog should increase (Bellegarda & Christof, 2016). Research shows that people speak over 15,000 words per day (Mehl, 2007). A child hears 20,000 words a day (Risley & Hart, 2006), adults probably more. These observations demonstrate the enormous potential of text data that, though currently ignored, will certainly be included in future comprehensive liflogs.

3 Induction of Semantic Facets in Textual Lifelog Data

In order for lifelogging to be useful as a tool for measuring cognitive activity, we will have to be able to classify a user’s daily cognitive activity through natural language processing of text that they create or consume (whether it come from reading, writing, seeing, speaking or hearing). It is easy to perform word-based index textual data; it is harder to organize it into cognitive activities. And though much work has been done for capturing episodic activity (Gurrin, Smeaton, & Aiden, 2014), liflogs do not typically capture or store cognitive activities and this will have to change (Wang, Peng, & Smeaton (2011).

Responding to this challenge of enriching diverse and massive, personal lifelog data, we have designed a private, personal search platform for capturing and classifying semantically classified cognitive data from a person’s digital interactions (source (i) above, the other three sources will be treated in future versions). In their private space, a user provides credentials for their personal data sources: email and social apps, as well as quantified-self apps. This diverse data is fetched and annotated using topic vocabularies in the form of taxonomies, that the user has chosen as representing their interests. The process of inducing these taxonomies is explained in Grefenstette (2015), and in Grefenstette and Muchemi (2015 and 2016). Search facets generated from these taxonomies facilitate semantic categorising and browsing of user-generated or user-consumed data. They could also help to measure the amount of text and time devoted to certain topics, as well as the amount of topic-specific vocabulary encountered. Suppose, for example, that a student is taking a “Managerial Accounting” course. One would expect their daily activity during that period to include some reading, hearing, browsing, and speaking about topics in this field. With our taxonomy induction technique we can automatically generate a domain taxonomy such as:

```
managerial_accounting>costs>variable costs>sales volume
managerial_accounting>costs>variable costs>selling price
managerial_accounting>costs>variable costs>transfer price
managerial_accounting>financing
managerial_accounting>financing>activity-based costs
managerial_accounting>financing>business decision
managerial_accounting>financing>financial accounting
managerial_accounting>financing>financial accounting>bookkeeping
```

**Figure 1. An Example of the Induced taxonomy for the topic “Managerial Accounting”**

This taxonomy includes a rich vocabulary related to the domain (..., financial ratios, financial report, financial reporting, financial reports, financials, financial statement, financial statement analysis, financial statements, find results, fixed cost, fixed costs, gaap, garrison, general accepted accounting principles, generally accepted accounting principles, graduate certificate, historical cost, historical costs, ...) that can be used to annotate lifelog entries as belonging to this topic, once the topic taxonomy is activated by the user. In our ongoing work we have tested our induced taxonomies to successfully distinguish topics in text sources such Reddit comments. We have created hundred of taxonomies for personal activities such as hobbies and illnesses. The results will be publicly available as soon as our experiments are complete, but in the meantime, it seems that it is feasible to easily create a large number of targeted vocabularies, and that these vocabularies can be used to classify daily activity into domains which can be used to measure the actual cognitive activity of future lifeloggers, just as quantified self tools can be used to measure physical activities today.

4 Conclusion

The partitioning of text-based lifelog data using domain taxonomies can facilitate analyzing of lifelog and classifying activity retrospectively. Though some attempts at developing systems that allow manual classification of lifelogged activity have been proposed (MyLifeBits (Gemmell, Bell & Lueeder, 2004), LifeLog (Kiyoharu, et al. 2004), Stuff I’veSeen (Dumais, et al. 2003), PERSONE (Kim, et al. 2006) Personal Data Prototype (Teraoka, 2012)), we feel that activity annotation must be an automatic and passive process. Loggerman (Hinbarji, et al., 2016) is a recent system that allows automatic logging of a person’s typing and app use. This is a good start, but we believe each piece of information that a user
generates or consumes must also be semantically classified and annotated. Annotating a person’s cognitive activity will allow the user, and anyone that the user shares their data with, to judge whether time spent learning is sufficient, and pooling the results of users will provide an additional dimension for improving directed and self-directed learning.

5 References


A ColourMap Visualisation of Lifelog Data

Aaron Duane, Moohamad Hinbarji, Rami Albatal, Cathal Gurrin
Insight Centre for Data Analytics,
Dublin City University, Ireland

Abstract
In this work, we motivate and present the EyeAware framework for lifelog data storage, analysis and access. We introduce the ColourMap interface to a visual lifelog and described how it operates and explain the next steps in this development process.

Keywords: lifelogging; test collection; personal data

Citation: Editor will add citation

Copyright: Copyright is held by the authors.

Acknowledgements: This research was supported by Science Foundation Ireland under grant number SFI/12/RC/2289.

Contact: aaron.duane@dcu.ie

1 Introduction
Lifelogging, which is the continual passive capture of life experience data in a digital archive is fast becoming a popular activity. Initially starting with Quantified Self applications, individuals are now exploring new ways of passively capturing a detailed trace of lifelog data. The reasons for this can be many-fold, such as personal wellness (Meyer et al., 2015), supporting episodic memory (Hodges et al., 2006) using wearable cameras, or even more recently, supporting semantic human memory by sensing applications that capture what a person reads and writes (Hinbarji et al., 2016).

Naturally such activities create vast archives of data, more than can be reasonably managed by the individual. Hence there have been initial efforts at organizing lifelog data to support multiple use-cases (Gurrin et al., 2014), guided by suggestions of possible applications (Sellen & Whittaker, 2010) which are referred to as the 5Rs of memory access. In this work we explore a novel lifelog visualization system for passive wearable camera data that supports an individual in reflecting on past activities upon both an overview (day and week) level and at a detailed ‘what did I do in that minute’ level. The contribution of this work is the description of the ColourMap visualization of visual lifelog data.

2 ColourMaps of Lifelog Data
Many lifeloggers are equipped with wearable cameras which can take hundreds of photographs passively throughout the day, from the viewpoint of the wearer. Typical wearable cameras cap capture up to 4,000 images in any given day; too many for manual organisation. Therefore a specialised lifelog data organisation system would be required that can store, analyse, organise and provide access facilities to the lifelog data.

The EyeAware framework is one such framework, that was designed to be a flexible and extensible server-based software infrastructure for managing large volumes of heterogeneous lifelog data, including visual data, locations, biometrics, and other sensor sources. The integration of data from such heterogeneous sources is achieved by utilising capture-time as evidence for linking together sensor streams. EyeAware integrates content analysis technologies, such as event detection, concept detection, sensor fusion (e.g. biometrics, location), face detection as underlying organisational tools for lifelogs.

EyeAware has been developed primarily to support browsing functionality with faceted filtering across axes such as visual concepts, biometric responses and location. This browsing-based access is supported by a number of interfaces that are all designed to support different user applications. One such interface is called the ColourMap interface, which was designed to explore and summarise lifelog data in a very simple yet visual way, at the two levels of overview and detailed.

Based on the concept of a heatmap, which is a graphical representation of data where the individual values contained in a matrix are represented as colours, the ColourMap takes this idea and extends it to support direct mapping of a colour value from related underlying visual data; hence the name ColourMap. The axes of the ColourMap can be flexible, and in the case presented here, it is a temporal distribution over one week of lifelog data.
The ColourMap interface utilises the minute (of time) as the basic unit of indexing and retrieval. Every day is segmented into 1,440 segments, corresponding to each minute in the 24 hour period. In total, seven days of data is displayed on every screen, with that limit based on the capabilities of the web browser, as can be seen in Figure 1. Although every minute segment can represent any type of data (e.g. stress level or heart rate, as in other EyeAware interfaces), for the ColourMap, the first image from any minute is used to extract the single most prominent colour from that image. This colour is then used to represent that minute in the interface. If no image is captured during a minute, the minute is represented by a blank white colour.

The minute segments for each day are arranged into seven columns which expand evenly across the width of the screen. Within each column the minutes move from left to right and from top to bottom; so the first minute of the day (00:00) is located in the top-left of the column and the last minute of the day (23:59) is located in the bottom-right of the column. On the left of the interface the Y axis is divided into 24 segments corresponding to each hour of the day to improve summary precision and coherence.

The minute segments act as access points into the lifelog, and selecting any one brings forth a thumbnail of the underlying image for about five seconds, as shown in Figure 2. If this is not interacted with, it disappears. However clicking on this thumbnail will expose the full resolution image to the user for them to review.
In the first iteration of the interface, the average colour was extracted from an image colour histogram. This process almost always resulted in a variation of grey as a representation. This monochromatic summary was not found to be insightful to users of the system. As an alternative, the idea of colour prominence was explored. In this context we refer to colour prominence as the most commonly occurring colour that contrasts most with the other colours inside an image. Although it is yet to be experimentally validated, we propose that this allows for the extraction of a more meaningful and aesthetically pleasing colour representation of an individual minute in which objects attracting attention (e.g. a colourful postbox, a colourful coat, the unique colour signature of a laptop screen) would be considered to be the predominant colour of an image.

Although we have not yet done a user study, our initial understanding is that the ColourMap interface provides a means for the data owner to reflect on life activities throughout the past week. The colours that occur in the life of the individual can be meaningful to that individual and provide ‘at a glance’ reflection opportunities over 20-30,000 lifelog images.

3 Conclusion
In this work, we presented the ColourMap interface to a visual lifelog and described how it operated. Thus far, one lifelogger has uploaded 1 million images to the ColourMap interface. The next steps in this research will be to engage in a real-world evaluation of the usefulness of the ColourMap concept as a means of an individual accessing and browsing their lifelog.

4 References


Hinbarji et al., (2016). LoggerMan, a comprehensive logging and visualisation tool to capture computer usage. In: 22st International Conference on MultiMedia Modelling (MMM 2016), 4-6 Jan, 2016, Miami, FA

NTCIR12-Lifelog, A Test Collection to Support Collaborative Benchmarking of Lifelog Retrieval Systems

Liting Zhou¹, Hideo Joho², Frank Hopfgartner³, Rami Albatai¹, Cathal Gurrin¹
¹Insight Centre for Data Analytics at Dublin City University, Ireland.
²Tsukuba University, Japan
³University of Glasgow, United Kingdom

Abstract
Test collections have a long history of supporting repeatable and comparable evaluation in Information Retrieval. In this paper we describe NTCIR12-Lifelog, a first lifelog test collection to support comparative benchmarking of lifelog retrieval systems. We provide an overview summary of the test collection and motivate use-cases of its application. This test collection is being employed at NTCIR-12 to evaluate approaches to Lifelog data search and organisation.

Keywords: lifelogging; test collection; privacy-by-design; personal data
Citation: Editor will add citation
Copyright: Copyright is held by the authors.
Acknowledgements: This research was supported by Science Foundation Ireland under grant number SFI/12/RC/2289.
Contact: becky.zhau@dcu.ie

1 Introduction
For many years, test collections have being used to support repeatable and comparative evaluation in Information Retrieval and related fields. The earliest example is the pioneering Cranfield collection of 1,398 abstracts of aerodynamics journal articles, a set of 225 queries, and exhaustive relevance judgments of all (query, document) pairs, which was gathered in the late 1950s to support early experimentation into Information Retrieval (Van Rijsbergen & Sparck Jones, 1972). Since then the field has embraced the concept of test collections for reasons such as:

• Lowering the cost of experimentation by removing the need for researchers to gather documents and construct relevance judgements, and;
• Helping to ensure that experiments are both repeatable and comparable across sites and time.

Consequently, the generation of test collections has been a primary focus of researchers in all aspects of Information Retrieval and related disciplines. The process has been formalised under the umbrella of a number of organisations such as TREC in the US, CLEF in Europe and NTCIR in Asia.

One aspect of Information Retrieval that has been gathering increasing attention in recent years is the concept of lifelogging. Lifelogging is defined as “a form of pervasive computing, consisting of a unified digital record of the totality of an individual’s experiences, captured multi-modally through digital sensors and stored permanently as a personal multimedia archive” (Gurrin, Smeaton & Doherty, 2014). Lifelogging typically generates multimedia archives of life-experience data in an enormous (potentially multi-decade) lifelog. Such a lifelog needs to be organised and searchable to be valuable to the lifelogger. Hence there have been calls for a test collection of lifelog data. The contribution of this paper is an overview description of the NTCIR12-Lifelog test collection.

2 A Description of the NTCIR-12 Lifelogging Test Collection
The NTCIR12-Lifelog test collection was created in 2015 as a multimodal dataset of three months of real-world lifelog data. Accompanying this dataset was a set of topics and relevance judgements supporting ad-hoc style retrieval.

The NTCIR Lifelog dataset consists of a subset of a real-world lifelog from three lifeloggers for a period of about one month each. The lifeloggers were asked to gather data during as much of their waking hours as possible, and typically consisted of all-day data. The data consists of a large collection of 88,124
wearable camera images (from the OMG Autogapher camera) and anonymised via the process previously described. An example of such images shown in Figure 1. The dataset also contained It also contains semantic locations (e.g. Starbucks cafe, McDonalds restaurant, home, work), physical activities (e.g. walking, transport, cycling) and an XML description of this data.

<table>
<thead>
<tr>
<th>Component</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of Lifeloggers</td>
<td>3</td>
</tr>
<tr>
<td>Size of the Collection (Days)</td>
<td>87 days</td>
</tr>
<tr>
<td>Size of the Collection (Images)</td>
<td>88,124 images</td>
</tr>
<tr>
<td>Size of the Collection (GB)</td>
<td>18.18 GB</td>
</tr>
<tr>
<td>Size of the Collection (Number of Long-stay Semantic Locations)</td>
<td>130 locations</td>
</tr>
<tr>
<td>Size of the Collection (Visual Concept Metadata)</td>
<td>840 MB</td>
</tr>
<tr>
<td>Number of LSAT (Ad-hoc) Topics</td>
<td>48</td>
</tr>
<tr>
<td>Number of Insight Topics</td>
<td>10</td>
</tr>
</tbody>
</table>

Table 1. A Basic Analysis of the NTCIR12-Lifelog Test Collection

Given the fact that lifelog data is typically visual in nature, the output of the CAFFE CNN-based visual concept detector (Jia et al, 2014) was included in the test collection as additional metadata. This classifier provided labels and probabilities of occurrence for 1,000 objects in every image. A summary of the test collection is shown in Table 1 and sample images in Figure 1. The data was roughly equally distributed per lifelogger.

Figure 1. Sample Wearable Camera Images from the NTCIR12-Lifelog Test Collection

Aside from the data, the test collection includes a set of topics (queries) that are representative of the real-world information needs of lifeloggers (Sellen & Whittaker, 2010). There are 48 ad-hoc search topics representing the challenge of Retrieval from memories. These topics were generated by the three lifeloggers and represent real information needs that they would have had for their lifelog data. The relevance judgements for NTCIR12-Lifelog were manual (non-pooled) relevance judgements, and were generated by the lifeloggers and data organisers for all 48 LSAT topics. These are used to compare the participant submissions and form the third component of the NTCIR Lifelog test collection that supports comparative and repeatable experimentation into organisation and search of lifelog archives.

3 References

Lifelogging by Senior Citizens in a Highly Ageing Society: A Pilot Study

H. Joho$^{1,2}$, M. Matsubara$^{1,2}$, N. Uda$^1$, K. Donkai$^{1,2}$, and C. Mizoue$^{1,2}$

$^1$Faculty of Library, Information and Media Science, University of Tsukuba, Japan
$^2$Research Center for Knowledge Communities, University of Tsukuba, Japan

Abstract

Many societies are rapidly ageing. In such a highly ageing society, one of the challenges is to establish metrics, measures and analytical approaches for better understanding diverse characteristics of senior citizens’ life. Lifelogging technologies can play an important role to address this problem, since it is now possible to gather long-term quantifiable self-data using advanced wearable devices and mobile applications. This paper reports our findings of a pilot study where four participants were involved in their lifelogging exercise using light-weight methods. The results suggest that our approach can capture diverse behavioural patterns of senior citizens without physical or mental burden.

Keywords: lifelogging; senior citizen; aging society

Citation: Editor will add citation.

Copyright: Copyright is held by the authors.

Acknowledgements: This work was partly supported by the JSPS 2014 Topic-Setting Program to Advance Cutting-Edge Humanities and Social Sciences Research Area Cultivation Project. Any opinions, findings, and conclusions described here are the authors and do not necessarily reflect those of the sponsors.

Contact: hideo@slis.tsukuba.ac.jp

1 Introduction

Many societies are rapidly ageing. United Nation’s World Population Aging 2015 Report (UN, 2015) states that “Between 2015 and 2030, the number of people in the world aged 60 years or over is projected to grow by 56 per cent, from 901 million to 1.4 billion” (p.2). Japan, for example, has already reached to the level where 26.9% of population is over 65 years old (Statistics Bureau, 2016). In such a highly ageing society, one of the challenges is to establish metrics, measures and analytical approaches for better understanding diverse characteristics of senior citizens’ life. However, World Health Organization states that “the current metrics and methods used in the field of ageing are limited, preventing a sound understanding of key aspects of Healthy Ageing” (WHO, 2015, p.221-222).

Lifelogging technologies can play an important role to address this problem, since it is now possible to gather long-term quantifiable self-data using advanced wearable devices and mobile applications. Moreover, effective analyses and visualisation of lifelog data can allow senior citizens to increase their ability of meta-cognition and reflection of their behavioural patterns (Schön, 1983; Hacker, Dunlosky, & Graesser, 1998). Such reflective learning is expected to enable them to set feasible goals for improving their well-being.

However, we have limited understanding of effective design and implementation of lifelogging in older age. This paper reports our findings of a pilot study where four participants were involved in their lifelogging exercise. The objectives of our pilot study was to examine the feasibility of our data collection approach, and to obtain initial observations of behavioural diversity.

2 Data collection approach

Lifelogging technologies can range from proactive intensive data collection to light-weight method (Gurrin, Smeaton, & Doherty, 2014). Since we did not have good understanding of senior citizens’ general capability for lifelogging, we decided to start with a light-weight method. More specifically, we asked participants to carry a 4-inch smartphone with 3G Internet connection whenever they went out.

Healthy Ageing is defined as “the process of developing and maintaining the functional ability that enables well-being in older age.” (WHO, 2015, p.28).
Figure 1: Moves smartphone application. Taken and modified from http://moves-app.com/.

(a) Participant A
(b) Participant B
(c) Participant C
(d) Participant D

Figure 2: Visited places of four participants.

Moves\(^2\), a GPS based location tracking application, was installed to the smartphone to record participants’ moving behaviour. As shown in Figure 1, Moves visualises the timeline of moving data along with transportation methods. The point of interest (POS) data is obtained from various sources in the application, so that participants manually determined the places they visited. However, the grounding of visited locations to POS was optional in our study. The collected data can be exported into multiple formats such as CSV, JSON, GPX, KML, or ICS, as well as in multiple intervals such as daily, weekly, or monthly. These data were first presented to participants for a data screening process. In the data screening process, participants had opportunity to review the recorded data and to remove any data that they did not wish to be analysed by the experimenters.

To control the level of background information, we recruited participants (two females and two males, age between 60 and 68) who were members of a family and their relatives, all living in the same city in Japan. This allowed us to observe the behavioural diversity of four participants even though the sample size was limited. We first presented an information sheet that explained the aim of the experiment and overview of the tasks involved in the experiment. Participants were then asked to sign a consent form if they agreed to participate.

The first session with participants included the hands-on workshop of smartphone (how to start, how to

\(^2\)http://moves-app.com/ (Visited: 24/02/2016)
tap, how to input letters, how to charge, etc.), and Moves application (how to start, how to view the result, how to ground POS, etc.). After the first session, we had a monthly meeting to gain feedback from participants. We started the experiment in January 2016, and the following results are based on the first three weeks of the data.

### 3 Initial result

The three weeks of location data collected by participants in January 2016 were visualised on the map as shown in Figure 2. The location data show a clear pattern of individual participants. For example, Participant A’s geographic range of visited places was the largest, followed by Participant D, B, and finally, C. Participant C’s data were mostly located within the city although the participant was quite active to visit different places within the city. Participant A and D had a similar pattern since they often traveled together in this particular time of period. Nevertheless, Participant A visited a wider range of places than Participant D.

Moving methods and their durations also showed clear differences among participants. Figure 3 show the duration of transport (Car, bus, and train) and walking of participants across three weeks. First, the duration of transport largely varied over three weeks within the same participant data (e.g., Participant A and B). Second, Participant C had far more walking exercise than other participants. This is an interesting contrast to the moving area of Participant C, which was the smallest among participants.

This highlights the importance of collecting and visualising multiple sources in lifelogging to accurately picture one’s behaviour. For example, one might think that Participant C is not an active person since the moving area is small. However, the walking duration data indicates that Participant C had a regular walking habit than other participants. This also suggests that our data collection methods can gain many insights into senior citizens’ behavioural patterns.

After a month of data collection, we had an interim interview with participants. They all reported that carrying the smartphone was very easy and no hassle at all. They also enjoyed looking at their past activities on the phone.

### 4 Conclusion and future work

Our ongoing project has been exploring ways to capture senior citizens’ behaviour patterns using lifelogging devices. Our initial result indicated that a smartphone with GPS tracking application was a good start to identify diverse individual differences. We plan to increase the size of participants in the coming year of the project, based on the findings of this pilot study.

### References


Lessons Learned from 8 Months of Lifelogging
Case of 2nd Year Undergraduate Students at the University of Tsukuba, Japan

N. Nagaoka¹, Y. Furuhashi¹, N. Minei¹, and H. Joho²
¹College of Knowledge and Library Sciences, School of Informatics, University of Tsukuba, Japan
²Research Center for Knowledge Communities, University of Tsukuba, Japan

Abstract
This paper reports the findings and lessons learned from 8 months of intensive lifelogging exercise carried out by three undergraduate students as part of a course at the College of Knowledge and Library Sciences, University of Tsukuba. We describe an overview of the course and tools used for lifelogging. Then, we discuss individual views on useful tools/data and not so useful tools/data based on the exercise. We also discuss challenges faced by students in terms of data collection, motivation, and application to behavioural changes.

Keywords: Lifelogging; undergraduate students; reflection

1 Introduction
Lifelogging is defined as “a form of pervasive computing, consisting of a unified digital record of the totality of an individual’s experiences, captured multi-modally through digital sensors and stored permanently as a personal multimedia archive” (Gurrin, Smeaton, & Doherty, 2014). Although reflective learning has been traditionally performed by professionals (Schön, 1983), the recent advance of wearable devices and other sensors started to enable a wider range of people to perform data-oriented reflective learning via lifelogging. However, reporting of lifelogging exercise performed by undergraduate students is still limited. This paper reports findings and lessons learned from 8 months of intensive lifelogging activities carried out by three Second Year undergraduate students (who are also the authors of this paper) as part of a class in the College of Knowledge and Library Sciences (CKLS), University Tsukuba, Japan.

1.1 Overview of lifelogging class
Thematic Studies are one of the characteristic classes in CKLS where lecturers can design an one-year long learning program so that students can gain a deeper insight into a given subject compared to conventional single term classes. Thematic Studies I-4 (referred to as the class in this paper) was specifically designed to allow students to carry out a long-term lifelogging activities using wearable devices and activity monitoring tools, to better understand their behaviour and life based on the analyses of quantified self data, and to set quantifiable goals to improve their life. It was also designed to offer opportunities to learn the strengths and limitations of lifelog data, and gain insight into the privacy issue of lifelog data based on their own experience. The lifelog class is a year-long class started in April 2015. Data collection began in May 2015. The following report is based on 8 months of data collection until December 2015.

1.2 Data collection history
There were seven categories of data in the lifelog class, as shown in Figure 1. They were sleeping, location (including activity), photos, diary, food, money, and sentiment. The sentiment data were semi-manually extracted from diaries. Since students needed some time to get used to data collection devices and tools, a
new data category was added gradually over the course of the class. Also, we added a new category based on interests emerged from students during the class (Sentiment is one such example).

1.3 Lifelogging tools

**Mi-Band** A wrist band activity tracker. It can record steps, rem and non-rem sleep, and other data (Xiaomi, 2016). We used it for sleeping data collection.

**Moves** GPS-based location tracking application (ProtoGeo Oy, 2016). It can record steps, cycles, transportations, point of interests. We used it for location and activity data collection.

**Photos** Students were asked to take at least one photo a day using their smartphone.

**Diary** Students were asked to make reflective diaries every week at the class. It mainly described their activities and events happened in day-to-day life.

**FoodLog** A smartphone app to keep a record of time and contents by taking a photo of the foods students had (foo.log, Inc., 2016).

**Zaim** A smartphone app to keep a record of expenses by taking a photo of receipts (Zaim, Inc., 2016).

**Sentiment** This was extracted by manually labelling frequently occurring keywords in the diaries. Mecab (Kudo, Yamamoto, & Matsumoto, 2004) was used for morphological processing and POS tagging of Japanese texts.

2 Lifelog Data Analyses

In this section, we report some of the highlights of lifelog data analyses conducted by three students.

2.1 Eating places and visited places (Student A)

Figure 2 shows an eating pattern of Student A from August to December. As can be seen, the student used convenient stores at least 10 times in every month to prepare for meals. Also, the total number of meals is less than 60, which is much lower than the expected value of 90 (30 days x 3 meals). As a result, the student came to realise that her eating habits were far from what was ideal.

Figure 3 compares the percentages of visited places over the months. The data was collected by Moves. As can be seen, holiday months (August and September) have different patterns than academic term months (June, July, October, November, and December). An unexpected data was the relatively large proportion of Others in November. The student often had an impression that she was an indoor person, but the data made her realise that she went out more often than what she remembered.
2.2 Walking distances and sleeping patterns (Student B)

Figure 4 shows the distance of walking recorded by Student B from May to December. If we take an average distance of two academic terms periods (Spring: May to July, Autumn: Oct to Dec), they are similar (4.6km vs. 4.2km). However, the figure shows that there are more 0 distance days in Autumn term than Spring term. Again, this was not something the student expected or realised. The student suspected that a lower temperature can be one factor for not going out in Autumn.

Sleeping pattern was something that students never recorded and analysed before. There are two noticeable trends from the sleep data (Figure 5). One is that there was more deep sleeping during academic term periods. However, a clearer trend was irregular sleeping time during the summer holiday period. This was someone expected but the magnitude of the difference exceeded the expectation of the student.

2.3 Meals patterns (Student C)

Figure 6 shows the distribution of foods and dishes taken by Student C during meals from July to December. In the last two months, there was a large increase of Udon noodle and Miso soup. The student recalled that this was due to attempt of cooking for meals. Although the student tried to cook meals as many as possible, it often ended up easy choices such as noodles and soups. However, reflection of the past food log data had a positive effect on an increase of vegetables in December.

3 Lessons learned from the exercise

This section discusses our findings and lessons to be learned from our intensive 8 months of lifelogging. In particular, we discuss individual views of useful lifelog tools and not so useful ones, and challenges in
3.1 Useful tools and not so useful one

**Student A**  Student A found the data of visited places, meals, and keywords extracted from diaries useful for identifying what she did, in where, with who information. Such information was useful for her to remember major events of a day for reflection. In particular, she found meals information particularly useful. This is because it also related to a rich set of other contexts such as how she prepared it, how she ate with, and where she had. This motivated her to look back her lifestyle and set a new goal for better eating patterns.

On the other hand, moving distances and photos were less useful since a similar set of information can be inferred from the visited places and keywords as described above. They could be useful for those who have regular exercise practices. The money information helped Student A to understand her expenses, but a new goal has not been set yet.

**Student B**  Student B found the expense log (Figure 7) useful since it clearly showed the change of expense categories over the months. For example, August showed a high proportion of Leisure category since she went out frequently. November and December have a very different expense pattern, which however, matched the patterns of photo categories and location data. All in all, monitoring expenses motivated her to control how she spent money.

Keywords extracted from the diary were also useful for reflecting major events of a month, since frequently appearing words indicated clear differences among months. For example, term months had school-related words, while exam periods frequently mentioned words related to those topics. Similarly, there were many travel-related words in the summer holiday season. Photos had an unexpected way of usefulness where days without major activities did not have a photo. Therefore, the lack of photos in the lifelog informed her that it was a quiet day.

On the other hand, food log was not useful for Student B since she lives with her family and did not frequently cook by herself. Therefore, there was no strong patterns in her meal patterns, and thus, it did not lead to an effective goal setting for improving her eating quality.

**Student C**  Student C found the actions of recording her lifelog impacting her behaviour more than collected data. For example, the food log required her to take a photo of every meal that she had. This motivated her to have three meals a day. Photos of meals also enabled her to identify the days where she did not cook herself. The expense log also required her to take a photo of receipts at shops and restaurants. However, since taking a photo of receipts was troublesome, she started to avoid going to shops more than necessarily. This saved her expenses.

Taking general photos also made her realise new insight. Student C was not a fan of photo-shooting before. However, since she started to record her lifelog using photos, there were more occasions of taking photos with her friends or visited places. This led her to share with other people, and to increase a level of communication.
On the other hand, Student C found it frustrating when the devices failed to capture accurate data of her activities. For example, activity tracker was useful to capture her sleeping pattern and activity level. However, it cannot record her naps while it counted as awake time when she turned over in bed. Therefore, her sleeping data was not so reliable. She had a similar experience with location data where it recorded a place she did not visit, and cycling distance was very accurate. These data can mislead her reflection of past activities.

3.2 Challenges of lifelogging

Student A  Student A found it difficult to take a photo every day. There seems to be three factors. First, there were days where no significant things happen and she had no idea what to take a photo. Second, sometime she found it uncomfortable to take a photo in public. Third, Student A used a smartphone to take a photo and kept using the smartphone, which caused a loss in her task performance.

Student A also found it difficult to record negative aspects of her life. She tended to avoid recording negative things happened to her so that she could move on. However, this made it difficult to accurately reflect those days. On a related matter, this whole regular reflective activity seemed to increase the occasions that made her feel down. Initially, Student A expected to find clues to improve her life quality, but it also ended up facing the data that she did not want to see.

Student B  Some devices were not so accurate in their measurements, and thus, Student B had to check and correct the data frequently, which was a lot of work. Sleeping data and GPS data were some of those cases. Student B found it challenging to transfer the reflected insight from the lifelog data into a new habit. For example, her data showed that she spent most Sundays at home, and therefore, she decided to make a habit of going out as much as possible on Sundays. However, it did not last long. This means that simple reflection of her day-to-day activities is not sufficient for an effective goal setting that are rewarding.

At a later stage of our program, we started to extract sentiment words from our diaries. However, since we did not have such an intention at the beginning, the diary tended to record facts and events but not so many sentiment data. As a result, analyses of sentiment words were not so meaningful to differentiate month-to-month characteristics.

Student C  The challenge Student C faced was to minimise the unnecessarily use of smartphone which was boosted by the smartphone applications used for lifelogging exercise. Also, some devices were not very accurate at recording the data. It was time-consuming to correct wrong data. This includes the GPS data and receipt scanning for expenses.

Another challenge was to transform the finding of lifelog data analysis into a behavioural change. For example, Moves data showed a decrease of moving distances in November. Therefore, Student C set a goal of cycling more than 60km a week, but it was difficult to achieve. However, frequent visit to a new friend enabled her to reach the goal before she noticed. This leads her to pay attention not only to the lifelog data but also to the change of her life in general.

4 Summary and outlook

This paper reported the exercise of 8 months lifelogging performed by three undergraduate students. As can be seen, the data patterns we observed varied across individual students. Moreover, the types of lifelog data and tools that students found it interesting and useful were also very different. As a result, findings from the data and their perspectives on lifelogging itself varied.

However, all students were able to collect a large amount of data in a long-term period, performed a range of analyses, found patterns from their daily behaviour, and finally, had a better understanding of themselves through reflective learning exercise. They also agreed that class mates were a significant factor to continue the exercise, which is in line with the findings of a community-based weight loss exercise study (Johnston, Rost, Miller-Kovach, Moreno, & Foreyt, 2013).
Another common finding was the difficulty in causing behavioural change using a new goal set by the analyses of their lifelog data. Students noticed that further efforts and creative actions were necessarily to achieve behavioural changes. This is one of the future directions for the students and this lifelog class.

Finally, this paper concludes with the descriptions of a summary and outlook provided by three students.

**Student A**  
Student A found it interesting to see that usefulness of data was not always intuitive. She thought photos would be useful but it was far more burden than she had imagined. On the other hand, meal data led her to reflect the locations and social data, which was not expected before she started. People are likely to have their own good source of lifelog data for their reflections.

It was her classmates who made it possible to keep logging her life. The weekly classes and classmates who shared the findings and difficulties were very important aspects of her motivation. Also, talking to other people about her lifelogging exercise sometimes influenced them to take a photo of meals. Therefore, lifelogging can also affect other people’s behaviour.

On the other hand, Student A would like to think again if lifelogging is really necessarily for her. This was partly because there were cases when she had to face negative events in her life, which did not always help. However, it was a great experience for her to learn that lifelogging is not only clarifying positive aspects of her life but also negative ones.

**Student B**  
The data showed that term periods and holiday seasons had very different life patterns. Also, photos and expenses clearly showed the characteristics of every month. However, the skills of an effective goal setting based on the patterns found in her lifelog data need to improve. Lifelogging will be more fun if she can gain a skill of setting a goal that was feasible and meaningful for improving her life quality. Student C expressed that she would like to continue the logging of expenses.

**Student C**  
Student C’s initial reaction to lifelogging was merely tracing what she did in the past days. However, it was gradually expanded to finding patterns, setting a goal or using the data for decision making. Furthermore, it led to some level of behavioural changes in her life. This chain of reactions was part of a fun of lifelogging.

In this class, we had classmates and we were able to share and compare our findings. It was interesting to see that everyone had a different opinion about the usefulness of data. Therefore, it is important to find own tools and data that are best to understand one’s behaviour. Student C expressed that she would like to continue the logging of meals, expenses, and photos.

**References**