Overview of NTCIR-12 Lifelog Task

Cathal Gurrin
Dublin City University
cathal.gurrin@dcu.ie

Hideo Joho
University of Tsukuba
hideo@slis.tsukuba.ac.jp

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

Liting Zhou
Dublin City University
becky.zhou@dcu.ie

Rami Albatal
Heystaks Technologies Ltd.
rami.albatal@heystaks.com

ABSTRACT

In this paper we review the NTCIR12-Lifelog pilot task, which ran at NTCIR-12. We outline the test collection employed, along with the tasks, the eight submissions and the findings from this pilot task. We finish by suggesting future plans for the task.

Subtasks

Lifelog Semantic Access Task (LSAT)
Lifelog Insight Task (LIT)

Keywords

lifelog, test collection, information retrieval, multimodal, evaluation

1. INTRODUCTION

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"* [5]. Lifelogging typically generates multimedia archives of life-experience data in an enormous (potentially multi-decade) lifelog. However, lifelogging has never been the subject of a rigorous comparative benchmarking exercise, even though there have been calls for a test collection of lifelog data [5, 9].

In this paper we describe the NTCIR12-Lifelog pilot task. We begin with a description of the requirements for the lifelog test collection, followed by a description of the test collection itself. We then describe the two sub-tasks that were organised for this pilot task, before outlining the eight submissions and the results of these submissions. Finally we outline plans for the next edition of the Lifelog task at NTCIR.

2. TASK OVERVIEW

This pilot lifelog task aims to begin the comparative evaluation of information access and retrieval systems operating over personal lifelog data. This task includes of two sub-tasks, both (or either) could have been participated in independently. The two sub-tasks were:

- Lifelog Semantic Access Task (LSAT) to explore search and retrieval from lifelogs, and
- Lifelog Insight Task (LIT) to explore knowledge mining and visualisation of lifelogs.

2.1 LSAT Task

The LSAT task was a typical known-item search task applied over lifelog data. In this subtask, the participants had to retrieve a number of specific moments in a lifelogger's life. We consider moments to be semantic events, or activities that happened at least once in the dataset. The task can best be compared to a known-item search task with one (or more) relevant items. Participants were allowed to undertake the LAST task in an interactive or automatic manner. For interactive submissions, a maximum of five minutes of search time was allowed per topic. The LSAT task included 48 search tasks, generated by the lifeloggers and guided by Kahneman's lifestyle activities [10]. In total, the LSAT task received 5 submissions.

2.2 LIT Task

The LIT task was exploratory in nature and the aim of this subtask was to gain insights into the lifelogger's daily life activities. It followed the idea of the Quantified Self movement that focuses on the visualization of knowledge mined from self-tracking data to provide "self-knowledge through numbers". Participants were requested to provide insights about the lifelog data that support the lifelogger in the act of reflecting upon the data, facilitate filtering and provide for efficient/effective means of visualisation of the data. The LIT task included ten information needs representing the idea that one would use a lifelog as a source for self-reflection. We did not intend to have an explicit evaluation for this task, rather we expected all participants to being their demonstrations or reflective output at the NTCIR conference.

3. DESCRIPTION OF THE LIFELOG TEST COLLECTION

The lifelog test collection described in this paper was developed for the Lifelog track of the NTCIR-12 evaluation forum. Prior to generating the test collection we defined a number of requirements for the collection:

- To be large enough to support a number of different retrieval tasks, but not so large as to discourage participation and use.
- To lower barriers-to-participation by including sufficient metadata, so that researchers interested in a broad range of applications, with a range of expertise, can utilise the test collection.

- To include appropriate and real-world lifelog data gathered in a conventional lifelogging situation.
- To consider the principles of privacy-by-design when creating the test collection, because personal sensor data (especially camera or audio data) carries privacy concerns.
- To include challenging and realistic topics representing real-world information needs, based on the experience of real-world lifeloggers.
- To include a set of relevance judgements that can be utilised both as a source of data for comparative evaluation as well as being later utilised as a source of training data for future experimentation.
- To be a reusable test collection that can support a number of years of research activities.

These requirements guided the test collection generation process.

3.1 Data Gathering Process

The data was gathered by lifeloggers who wore the lifelogging devices for most (or all) of the waking hours in the day. The lifeloggers gathered about one month of data each, giving a total of 79 days of data for the test collection. The data consisted of continuous lifelog images collected using an OMG Autographer wearable camera, as well as the output of the Moves lifelogging app (locations and physical movements). The OMG Autographer camera is worn on a lanyard around the neck, captures the daily activities of the wearer (from the wearer's viewpoint) and can operate for a full-day on a battery charge. This camera takes photos passively (i.e. without explicit user intervention) at about two images per minute. This camera is a later generation of the Microsoft Sensecam wearable camera [6] which was used in early lifelogging research. The Moves app is a smartphone app that automatically records user activity in terms of semantic locations and physical activities (e.g. waking, cycling, running, transport) by running in the background on a smartphone; as such it is also passively capturing data.

Following the data gathering process, there were a number of steps that were taken to ensure that test collection was both as realistic as possible, and took into account sensitivities associated with personal data:

- Temporal Alignment. It was important to ensure temporal alignment of the sensor data, given that it is from separate devices. It was necessary to check and resolve alignment problems (1-2 minutes) for one lifelogger by cross-referencing reported timestamps from the Autographer camera with clocks captured daily in the real-world.
- Data Filtering. Given the personal nature of lifelog data, it was necessary to allow the lifeloggers to remove any lifelog data that they may be unwilling to share. Following this, all images were reviewed by one individual with oversight of the entire collection to ensure that no potentially embarrassing or offensive images were concluded in the collection.
- Privacy Protection. Privacy-by-design [1] was one of the requirements for the test collection mentioned. Consequently, two steps were taken to ensure privacy of both the lifeloggers and any subjects or bystanders [4] captured in the lifelog data. Each recognisable face in every image was blurred in a manual process that took a number of weeks to accomplish. We had explored the potential of automated face detection and blurring, but there were a significant number of false positives and missed faces, when using off-the-shelf face detectors. In addition, every image was also resized down to 1024 x 768 resolution which had the effect of rendering any on-screen text virtually illegible. The Moves app naturally protects privacy of our lifeloggers by converting all locations from absolute locations to semantic locations, which resulted in sensitive absolute addresses being labeled as 'home' or 'work'.

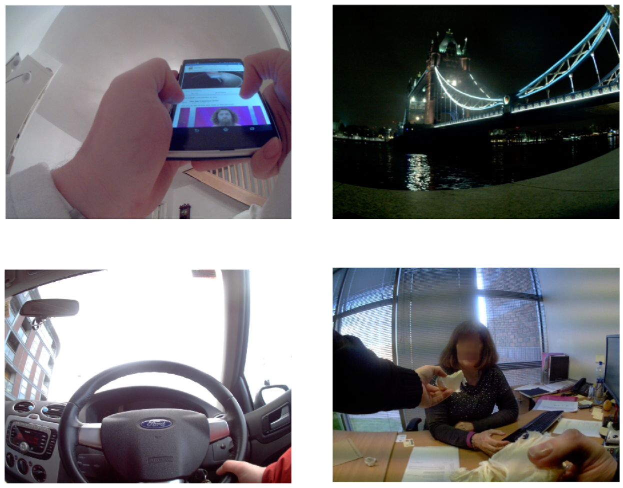

Figure 1: Examples of Wearable Camera Images from the Test Collection

3.2 Details of the Dataset

The NTCIR Lifelog test collection consists of data from three lifeloggers for a period of about one month each. The data consists of a large collection of wearable camera images (at about 2 per minute) as shown in Figure 1 and an XML description of the semantic locations (e.g. Starbucks cafe, McDonalds restaurant, home, work) and the physical activities of the user (e.g. walking, transport, cycling), of the lifelogger at a granularity of one minute.

Given the fact that lifelog data is typically visual in nature and in order to reduce the barriers-to-participation, the output of the CAFFE CNN-based visual concept detector [8] was included in the test collection as additional metadata. This classifier provided labels and probabilities of occurrence for 1,000 objects in every image. The accuracy of the CAFFE visual concept detector is very variable, and is representative of the current generation of off-the-shelf visual analytics tools.

A summary of the test collection is shown in Table 1. All three lifeloggers gathered (visual) data during most of the waking day, though User 1 wore the camera for longer days than the other two users, as shown in Figure 2, which also shows the distribution of topics per lifelogger. As can be seen, User 1 has more images and topics than Users' 2 and 3. User 1 wore the camera slightly longer and was engaged

Number of Lifeloggers	3
Size of the Collection (GB)	18.18GB
Size of the Collection (Images)	88,124 images
Size of the Collection (Locations)	130 locations
Size of the Collection (Visual Concepts)	825MB
Number of LSAT Topics	48
Number of LIT Topics	10

Table 1: Statistics of NTCIR-12 Lifelog Data

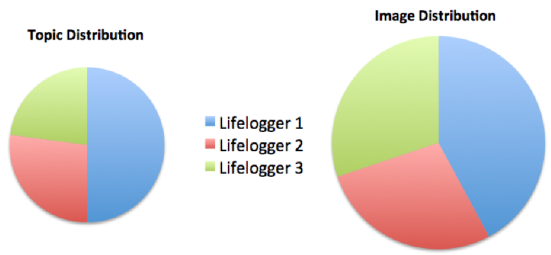

Figure 2: Distribution of Topics and Images per Lifelogger

TITLE: Tower Bridge
DESCRIPTION: Find the moment(s) when I was looking at Tower Bridge in London.
NARRATIVE: To be considered relevant, the full span of Tower Bridge must be visible. Moments of crossing the Tower Bridge or showing some subset of Tower Bridge are not considered relevant.

Figure 3: LSAT Topic Example

in a wider-range of activities than the other two users.

3.3 Topics

Aside from the data, the test collection includes a set of topics (queries) that are representative of the real-world information needs of lifeloggers and represent the *Retrieval* and *Reflection* reasons for accessing memories [14]. There are 48 ad-hoc search topics representing the challenge of *Retrieval* for the LSAT task, called the LSAT (Lifelog Semantic Access) Topics. These LSAT topics were evaluated in terms of traditional Information Retrieval effectiveness measurements such as Precision, Recall and NDCG. An example of an LSAT topic is included as Figure 3. For a full list of the topics see Table 2. In this table, the number of groundtruth relevant events are shown for each topic, as well the recall figures for the top performing automatic and interactive LSAT topics.

Additionally, there were ten insight topics representing the challenge of supporting *Reflection* from memories. These were called LIT (Lifelog Insight) Topics and are not evaluated in a traditional sense. Participants were encouraged to prepare insights and demonstrate them directly to other participants at the NTCIR-12 Conference. An example of an LIT topic is included as Figure 4.:

3.4 Relevance Judgements

Manual (non-pooled) relevance judgements were generated manually 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.

TITLE: Early Morning Commute
DESCRIPTION: Early Provide insights on the methods of, and duration, each lifelogger spends commuting to work.
NARRATIVE: Commuting to work or university, via whatever means, is relevant. Commuting to a meeting in a location that is not the user's normal place of work is also relevant if it could be considered to be a morning commute to work. Commuting home is not relevant. General travelling is not relevant.

Figure 4: LIT Topic Example

4. PARTICIPANTS AND SUBMISSIONS

In total, five participants submitted to the LSAT task and three participants submitted to the LIT task (at time of writing). This was from a total of sixteen groups who signed up to take part in the Lifelog Task at NTCIR-12. Given that the LIT subtask is not evaluated in the traditional sense, we report only on the LSAT subtask runs that we received, organised by automatic and interactive tasks.

4.1 LSAT Task

For the LSAT task, each participant took a different approach to generating submissions. Although the LSAT task supported both automatic and interactive submissions, four of the five participating groups submitted automated runs only and only one group developed an interactive retrieval system and submitted three runs generated by users in an interactive manner. We understand that this was because of the significant effort required to develop a user-friendly interactive retrieval engine. We firstly describe the automatic LSAT submissions.

4.2 Automatic LSAT Submissions

We now describe the four automatic LSAT submissions: *VTIR, USA.* The VTIR team (CBIA in Figure 5) identified that location was a very important component in the information retrieval process [15]. Thus, 3,000 images were randomly chosen from the dataset and manually labeled with a rich semantic location ontology (including office, home, kitchen, street, transportation, etc). This allowed for more than 80% accuracy in determining the locations. The visual concepts distributed with the collection were enriched by applying the WordNet Database to find the sets of cognitive synonyms. For each query, the location features were used as part of the retrieval process. After testing different models, BM25 provided the best performance; the parameters optimised and a optimal thresholds found and applied. Overall it was found that this approach worked well for some queries, but not for others.

IDEAS Institute for Information Industry, Taiwan. The Ideas team (III&CYUT in Figure 5) took a textual approach to retrieval [11]. As a baseline, a word distance measure between the provided CAFFE concepts and the keywords retrieved from lifelog tasks was examined using the word2vec model provided by Google. These pre-trained vectors trained on part of Google News dataset (about 100 billion words), and the model contains 300-dimensional vectors for 3 million words and phrases. This was extended by the application of the Stanford NLP parser on the task

Topic Title	Total Relevant	Recall (Automatic)	Recall (Interactive)
The Red Taxi	1	1	1
Photographing a Lake	2	2	N/A
Presenting/Lecturing	3	2	0
Tower Bridge	1	1	1
Driving a Rental Car	19	0	5
Attending a Lecture	1	0	0
Eating while in Conversation	1	1	1
On the Bus or Train	4	2	1
New Key	1	0	1
Having a Drink	2	0	1
Lost	1	1	N/A
Riding a Red Train	3	2	3
Man in a Burberry Coat	1	1	N/A
The Church	1	1	1
The Rugby Match	3	1	3
Costa Coffee	4	2	2
Antiques Store	3	3	3
Outdoor Computing	2	0	1
Building a Computer	14	1	1
Airbus A380	1	1	1
Shopping for a Bottle of Wine	1	1	N/A
ATM	3	3	1
Shopping For Fish	3	2	3
Repairing a Car Wheel	1	0	1
Cycling home	7	0	5
Happy Homework	1	0	N/A
Shopping	14	9	5
Informal Coffee Meeting	1	1	N/A
Lunchtime	8	4	0
In a Meeting	8	4	1
Bus to the Airport	1	1	1
A Movie on the Flight	1	1	0
The Metro	12	9	3
A Garden Chat with Dog	1	1	1
Lion at the Gate	1	1	1
Checking the Menu	3	0	0
The BirdâĂŹs Nest Stadium	3	2	1
Watching TV	21	0	2
Grocery Store	39	11	18
Strolling on the Deck	1	0	N/A
Eating on the Roadside	1	0	1
Writing	51	4	1
The Elevator	19	7	7
Car Repair	1	0	1
Drinking in a Pub	18	6	11
Barbershop	1	1	N/A
Lottery	1	1	N/A
Checkout	30	2	4

Table 2: Statistical Analysis of NTCIR-12 Lifelog Data

descriptions and sentiment analysis to identify negative feature keywords. A final submission utilised query expansion on every keyword in an attempt to enhance retrieval performance.

LIG-MRM, France. The LIG-MRM group (LIG-MIRM in Figure 5 and Automatic in Figure 6) focused on enhancing the performance of the visual concept detectors to be used for retrieval, and not relying on the CAFFE classifier output [12]. This was achieved using three techniques, including Dynamic Convolutional Neural Networks VGG (1000 dimensional features of ImageNet), a classification provided by their own MSVM on optimized data TRECVid (346 dimensional features), and a third approach utilising MSVM classification on VOC concepts (20 dimensional features). The images were also described by their respective annotated metadata (when present), such as place (e.g. "home", "work", etc.) and activity (e.g. "walking", "transport", "bus", etc.). When processing a topic Q, a mapping (string inclusion) of each word from Q into one or more visual concepts was performed. The score of each image was then computed

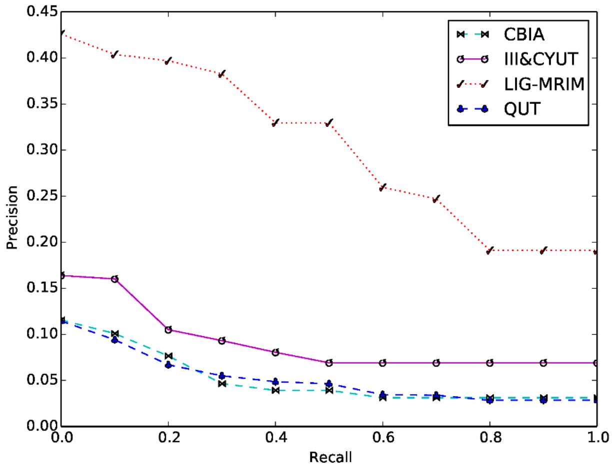

Figure 5: Comparing the best runs of the four automatic LSAT teams

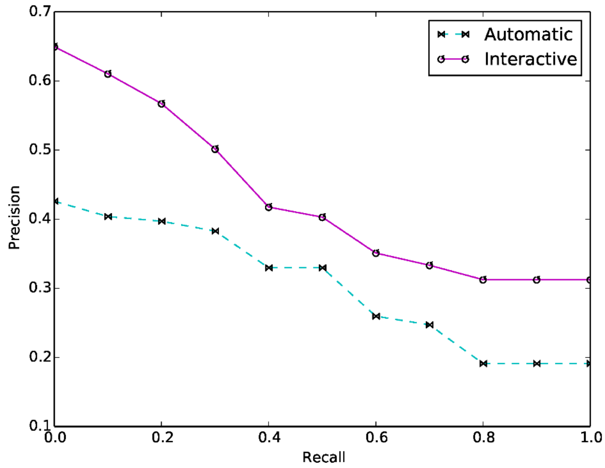

Figure 6: Comparing the best automated LSAT run to the best interactive LSAT run

as a fusion (linear combination) of the classifiers' scores for the mapped concepts of Q. Additionally, an image filtration process was utilised that selected images that meet certain location/activity criteria, when they existed.

QUT, Australia. The QUT group (QUT in Figure 5) took an approach to retrieval that utilised long, descriptive paragraphs of text to annotate the lifelog content, as opposed to the conventional tag-based approach [13]. These paragraphs then formed the basis for the retrieval technique. The process worked by temporally and visually clustering images, annotating a small number of images, and then spreading the annotations inter-cluster based on visual similarity of the clusters. Retrieval is then performed over the newly annotated content. Initial results suggest that the enhanced annotation process does not improve retrieval performance over a baseline of indexing the provided metadata and concepts.

University of Barcelona, Spain. The group from Barcelona (Interactive in Figure 6) were the only group to develop and run an interactive system [2]. Three runs were submitted, each of which was performed by four different users who ran a subset of the topics. The runs were distinguished by employing a semantic content-tagging tool and the inclusion of runs performed by either novice or expert users. The comparative performance of this interactive system on a topic-by-topic basis, when compared to the best performing automatic runs is shown in Table 2 (see the right-most two columns). When N/A is shown, it means that no moments were included in the submission for that topic (i.e. that the user using the interactive system could not find any potentially relevant moments for the given topic).

Although there was only one group participating in the interactive LAST task, it is possible to explore the result of this group in terms of the duration of search. When submitting interactive runs, groups were asked to include the number of seconds elapsed in the search process when that image was found by the interactive subject user. This allows for the comparison of the performance of an interactive system across different durations (from ten seconds to five minutes) as shown in Table 3, which shows the best performing run from the Barcelona group at five different durations of search.

As can be seen, the system performs significantly better when the user has more time to search and browse the collection, as would be expected.

4.2.1 Moments & Images as Submission Formats

A submitted run for the LSAT sub-task was in the form of a CSV file in the following format: *[topic id, image id (from within the relevant moment), seconds elapsed, belief score].* A moment was identified in the submission file as an image ID that answers the topic. If there are more than one sequential images that answer the topic (i.e. the moment is more than one image in duration), then any image from within that moment was acceptable. At evaluation time, each image ID is mapped into moments and evaluated at both the moment-level and the image-level. In this overview paper, we report on the moment-level evaluation. Both the moment-level and the image-level evaluations are included in the official results of the task, along with the mapping from image to moment (event).

4.3 LIT Task

For the LIT task, there were no submissions to be evaluated in the traditional manner; rather the LIT task was an exploratory task to explore a wide-range of options for generating insights from the lifelog data. Three groups took part in the LIT task:

Sakai Lab at Waseda University, Japan. The Sakai Lab utilised a prototype smartphone application called Sleepflower, which is designed to improve the sleep cycles of a group of users through a collaborative effort, hence the focus was on understanding the sleep cycles of the three lifeloggers represented by the dataset [7]. A flower metaphor is displayed on the smartphone screen to represent the current sleepiness of a particular user, along with similar metaphors for the other group members, in the hope of improving the lifestyles of the group as a whole. One significant limitation of the current prototype is that sleep hours and sleepiness data need to be entered manually; we are hoping to build a new prototype that semi-automatically collect lifelog data such as those provided by the NTCIR Lifelog task. As an

Time Elapsed	Num. of Relevant Moments Found	Number of Topics for which a Relevant Moment was Found
10s	7	3
30s	20	9
60s	32	15
120s	56	27
300s	94	34

Table 3: Interactive Run Comparison over Time

initial step towards this goal, the NTCIR Lifelog data was manually analysed from the viewpoint of individual sleeping habits and in our paper we discuss possible approaches to leveraging such data for the next version of Sleep-flower.

Toyohashi University, Japan. The group from Toyohashi examined repeated pattern discovery from lifelog image sequences, by applying a Spoken Term Discovery technique, which is an approach usually used to words from speech data [16]. A variant of Dynamic Time Warping was used in an experimental approach to extract extract meaningful patterns from the lifelog data. It is suggested that this could be a useful approach for insight generation from archives of lifelog data.

Dublin City University, Ireland. The submission from Dublin City University introduced an interactive lifelog interrogation system which allowed for manual interrogation of the lifelog dataset for the occurrence of visual concepts that were assumed to match the information needs [3]. The results of this manual interrogation were then used to generate insights and infographics for the provided topics.

5. LEARNINGS / FUTURE PLANS

This was the first collaborative benchmarking exercise for lifelog data. It attracted eight participants, four for the automatic LSAT search task, one for the interactive LSAT search task and three for the LIT task. We can summarise the learnings from this pilot task as follows:

- There is still no standardised approach to retrieval of lifelog data; each of the participants in the LSAT task took different approaches to retrieval. This suggests that the LSAT task is valuable to run again in future years.
- The dataset should contain more semantically rich data to support mare groups to take part. Such data should try to capture the semantics of daily life, and not just the activities.
- The supplied metadata and visual concepts should be at a higher-level of quality. The positive effect of higher-quality metadata can be seen in the results of the best-performing automatic LSAT run.
- The LSAT task is a valuable task, though effort should be made to encourage more interactive participants.
- The LSAT unit-of-retrieval being the moment/event required the mapping of submitted image IDs to moments that are defined by the coordinators in a manual process. This suggests that the unit of retrieval itself could become an interesting task, so we propose to include an event-segmentation task in the next running of the lifelog collaborative benchmarking exercise.

6. CONCLUSION

In this paper, we described the data and the activities from the first lifelog pilot-task at NTCIR. There were two sub-tasks, the LSAT known-item search task and the LIT data insights task. These task were undertaken by five and three participants respectively. The results attained by the participants showed that although many different approaches for automatic retrieval were applied, the one that appeared to work best was a computer-vision-based approach that augmented the provided CAFFE concept detector output with an enhanced concept detector based on a CNN-based model. In terms of interactive vs. automatic search, the interactive system performed better, as would be expected. The LSAT task is a valuable task and should continue, though perhaps joined by additional tasks, such as event segmentation. It is proposed that a bigger and more semantically rich dataset be employed for any future lifelog comparative benchmarking tasks.

Acknowledgements

We acknowledge the financial support of Science Foundation Ireland (SFI) under grant number SFI/12/RC/2289 and the input of the DCU ethics committee and the risk & compliance officer. We acknowledge financial support by the European Science Foundation via its Research Network Programme “Evaluating Information Access Systems”. The project has been reviewed and approved by the Research Ethics Committee of the College of Arts of University of Glasgow.

7. REFERENCES

[1] A. Cavoukian. Privacy by design: The 7 foundational principles. implementation and mapping of fair information practices. *Information and Privacy Commissioner of Ontario, Canada*, 2010.

[2] G. de Oliveira Barra, A. C. Ayala, M. BolaÃśos, M. Dimiccoli, M. Aghaei, M. CarnÃľ, X. Giro-I-Nieto, and P. Radeva. Lemore: A lifelog engine for moments retrieval at the ntcir-lifelog lsat task. In *Proceedings of NTCIR-12, Tokyo, Japan*, 2016.

[3] A. Duane, C. Gurrin, L. Zhou, and R. Gupta. Visual insights from personal lifelogs - insight at the ntcir-12 lifelog lit task. In *Proceedings of NTCIR-12, Tokyo, Japan*, 2016.

[4] C. Gurrin, R. Albatal, H. Joho, and K. Ishii. A privacy by design approach to lifelogging. In *Digital Enlightenment Yearbook 2014*, pages 49–73. IOS Press, 2014.

[5] C. Gurrin, A. F. Smeaton, and A. R. Doherty. Lifelogging: Personal big data. *Foundations and Trends in Information Retrieval*, 8(1):1–125, 2014.

[6] S. Hodges, L. Williams, E. Berry, S. Izadi, J. Srinivasan, A. Butler, G. Smyth, N. Kapur, and K. R. Wood. Sensecam: A retrospective memory aid. In *UbiComp 2006*, pages 177–193, 2006.

[7] S. Iijima and T. Sakai. Slll at the ntcir-12 lifelog task: Sleepflower and the lit subtask. In *Proceedings of NTCIR-12, Tokyo, Japan*, 2016.

[8] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J. Long, R. Girshick, S. Guadarrama, and T. Darrell. Caffe: Convolutional architecture for fast feature embedding. In *ACM MM*, pages 675–678, 2014.

[9] G. J. Jones, C. Gurrin, L. Kelly, D. Byrne, and Y. Chen. Information access tasks and evaluation for personal lifelogs. 2008.

[10] D. Kahneman, A. B. Krueger, D. A. Schkade, N. Schwarz, and A. A. Stone. A survey method for characterizing daily life experience: The day reconstruction method. *Science*, 306:1776–1780, 2004.

[11] H. L. Lin, T. C. Chiang, L.-P. Chen, and P.-C. Yang. Image searching by events with deep learning for ntcir-12 lifelog. In *Proceedings of NTCIR-12, Tokyo, Japan*, 2016.

[12] B. Safadi, P. Mulhem, G. Qul'not, and J.-P. Chevallet. Mrim-lig at ntcir lifelog semantic access task. In *Proceedings of NTCIR-12, Tokyo, Japan*, 2016.

[13] H. Scells, G. Zuccon, and K. Kitto. Qut at the ntcir lifelog semantic access task. In *Proceedings of NTCIR-12, Tokyo, Japan*, 2016.

[14] A. J. Sellen and S. Whittaker. Beyond total capture: A constructive critique of lifelogging. *Commun. ACM*, 53(5):70–77, May 2010.

[15] L. Xia, Y. Ma, and W. Fan. Vtir at the ntcir-12 2016 lifelog semantic access task. In *Proceedings of NTCIR-12, Tokyo, Japan*, 2016.

[16] K. Yamauchi and T. Akiba. Repeated event discovery from image sequences by using segmental dynamic time warping: experiment at the ntcir-12 lifelog task. In *Proceedings of NTCIR-12, Tokyo, Japan*, 2016.