Book Section:

Chapter 13
Experiments in Lifelog Organisation and Retrieval at NTCIR

Cathal Gurrin, Hideo Joho, Frank Hopfgartner, Liting Zhou, Rami Albatal, Graham Healy, and Duc-Tien Dang Nguyen

Abstract Lifelogging can be described as the process by which individuals use various software and hardware devices to gather large archives of multimodal personal data from multiple sources and store them in a personal data archive, called a lifelog. The Lifelog task at NTCIR was a comparative benchmarking exercise with the aim of encouraging research into the organisation and retrieval of data from multimodal lifelogs. The Lifelog task ran for over 4 years from NTCIR-12 until NTCIR-14 (2015.02–2019.06); it supported participants to submit to five subtasks, each tackling a different challenge related to lifelog retrieval. In this chapter, a motivation is given for the Lifelog task and a review of progress since NTCIR-12 is presented. Finally, the lessons learned and challenges within the domain of lifelog retrieval are presented.

C. Gurrin (✉) · L. Zhou · G. Healy
Dublin City University, Dublin, Ireland
e-mail: cathal.gurrin@dcu.ie

L. Zhou
e-mail: Liting.zhou@insight-centre.org

G. Healy
e-mail: graham.healy@dcu.ie

H. Joho
University of Tsukuba, Tsukuba, Japan
e-mail: hideo@slis.tsukuba.ac.jp

F. Hopfgartner
University of Sheffield, Sheffield, UK
e-mail: f.hopfgartner@sheffield.ac.uk

R. Albatal
SeekLayer Ltd, Dublin, Ireland
e-mail: ramibatal@gmail.com

D.-T. D. Nguyen
University of Bergen, Bergen, Norway
e-mail: ductien.dangnguyen@uib.no

13.1   Introduction

Recent advances in computing technology and wearable sensors mean that individuals are now in a position to log data about their lives on a continual basis, with little manual effort required. These data logs are often called lifelogs, and the process of gathering them is referred to as lifelogging. Lifelogging typically occurs in a passive manner (i.e. using sensors and not relying on human input). A commonly used definition of lifelogging is as *'a form of pervasive computing, consisting of a unified digital record of the totality of an individual's experiences, captured multimodally through digital sensors and stored permanently as a personal multimedia archive'* (Dodge and Kitchin 2007). Lifelogging can generate enormous (potentially multi-decade) archives that are too large for manual organisation. What sets lifelogging apart from conventional personal data organisation challenges (e.g. photos or emails) is the fact that lifelogs, being captured passively, are typically continuous in nature and are non-curated archives. Hence, these lifelogs pose a significant challenge for researchers to develop appropriate information organisation and retrieval approaches.

In the past 15 years, lifelogging has been receiving increasing research attention across a range of domains, including multimedia analytics, event-based computing, pervasive computing, information retrieval, as well as various application domains such as memory-science, wellness and epidemiological studies. A detailed overview of the early research activities on lifelogging is provided by (Gurrin et al. 2014b), and we refer the reader to that overview. Prior to NTCIR-12, there was no forum that could support a comparative evaluation of approaches to lifelog data organisation and retrieval, nor were the suitable datasets, nor was there even consensus on which of the many potential research challenges were the most important. The Lifelog task at NTCIR-12 was proposed because the organisers identified that lifelogging had potential to become a relatively commonplace activity, thereby necessitating the development of new approaches to multimodal personal data analytics and retrieval. New personal sensors were coming to market, such as wearable cameras, AR glasses, various forms of fitness trackers and so on. These were generating large multimodal archives for individuals yet, as with many new technologies, the required organisation tools had not been considered. It is the belief of the organisers that such vast archives of personal data require search and automatic annotation as fundamental underlying technologies upon which various other applications can be built; hence, the Lifelog task was proposed.

13.2   Related Activities

Lifelog data has been used in many domains as a source of multimodal data logging the real-world activities of one, or more, individuals. From prior research, we note that lifelogging tools were applied in the domains of long-term memory understanding (Milton et al. 2011), supporting human recollection (Barnard et al. 2011),

supporting human memory (Berry et al. 2007; Harvey et al. 2016), facilitating large-scale epidemiological studies in healthcare (Signal et al. 2017), lifestyle monitoring at the individual level (Nguyen et al. 2016; Wilson et al. 2018), behaviour analytics (Everson et al. 2019), diet/obesity analytics (Zhou et al. 2019), or for exploring societal issues such as privacy-related concerns (Hoyle et al. 2014). For many of these domains of application, the lifelog data was gathered and analysed by humans in order to draw conclusions for their research tasks.

In terms of actual functional retrieval systems for lifelog data, a number of early retrieval engines had been developed prior to NTCIR-12, such as the MyLifeBits system (Gemmell et al. 2002) or the Sensecam Browser (Lee et al. 2008). Both of these systems were browsing engines, rather than search engines, and relied on a database metaphor to support access. Subsequently, it was found that a faceted-multimodal search engine (even a simple one) was many times faster and more effective than browsing systems at finding known items from large lifelogs (Doherty et al. 2012), yet there were few search engines designed for lifelog data and no means of comparing their effectiveness. This means that prior to the Lifelog task at NTCIR-12, there were no comparative benchmarking activities and comparative and reproducible research on lifelogging was rather sparse. The main reason for this was the lack of publicly available lifelog datasets, which was due to the highly personal nature of lifelog data and the related requirement to guarantee people's privacy when releasing such datasets for widespread use.

The NTCIR-12 Lifelog pilot task (Gurrin et al. 2016) introduced the first shared test collection for lifelog data and attracted the first cohort of participants to, what was at the time, a very novel and challenging task. Since this initiative at NTCIR-12, there have been two related activities at alternative venues; one at ImageCLEF (Dang-Nguyen et al. 2017a, 2018) which focused on a series of image-retrieval and summarisation focused benchmarking initiatives since 2017, and the Lifelog Search Challenge (LSC) (Gurrin et al. 2019b) which was modelled on the successful Video Browser Showdown (Lokoc et al. 2018). The LSC encourages participants to develop interactive search engines for lifelog data and evaluate them in a public forum. The LSC has run at the annual ACM ICMR conference since 2018.

Specifically in relation to standalone retrieval efforts, early research on lifelog retrieval has focused on using *images* as unit of retrieval (e.g. Lee et al. 2008) with some early work in supporting user browsing these image collections (Doherty et al. 2011), or on the use of maps metadata, such as GPS locations, to organise content visually (Chowdhury et al. 2016). Once again, we refer the reader to (Gurrin et al. 2014b) for an overview of early efforts at lifelog search and retrieval. Significant efforts also went into the development of graphical user interfaces to visualise the data and also provide a positive user experience. Many good examples of interactive interfaces can be seen in the systems developed for the interactive Lifelog Search Challenge since 2018.

13.3   Lifelog Datasets Released at NTCIR

Over the course of the three most recent NTCIR workshops, the Lifelog task introduced three new datasets. The datasets were developed to represent a multimodal digital surrogate of the life activities of a number of individuals as they go about their daily lives, over an extended period of time (weeks or months). These datasets represented unprecedented data-rich archives for a number of individuals, pushing the boundaries of what was feasible to collect and distribute in an ethically and legally acceptable manner. Each dataset was gathered by either two or three lifeloggers, who wore/carried with them various lifelogging devices and gathered activity/biometric data for most (or all) of the waking hours in the day. The three datasets contained images from passive-capture wearable cameras as the core of each dataset. The passive-capture wearable camera was either clipped to clothing or worn on a lanyard around the neck, which captured images (from the wearer's viewpoint) and operated for 12–14 h per day (1,250–4,500 images per day—depending on capture frequency, camera type, or length of waking day). For examples of images captured by such wearable cameras, see Fig. 13.1. Additionally, mobile phone apps gathered contextual data such as location or physical movements and additional sensors (e.g. smartwatches or biometric-testing sensors) provided health and wellness data.
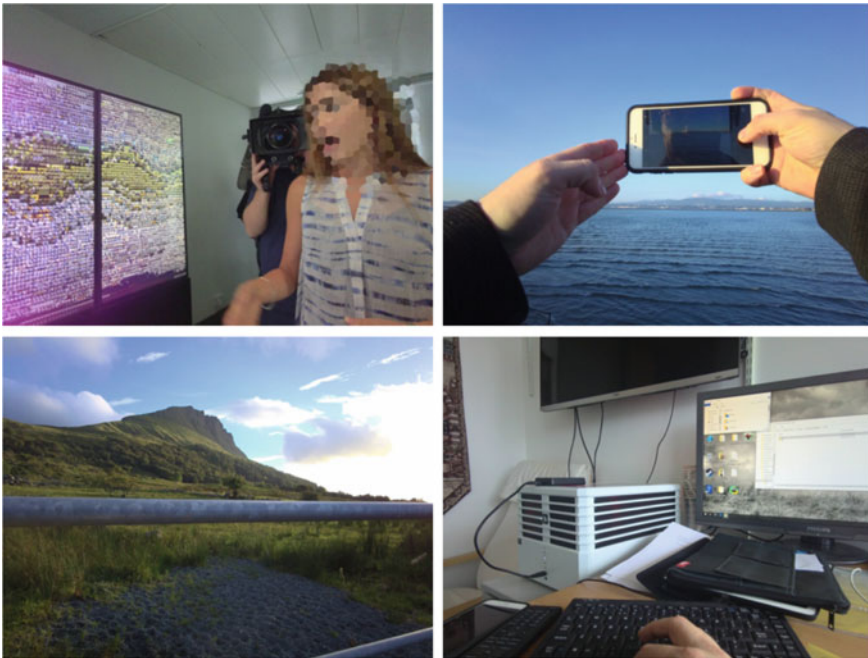

Fig. 13.1   Examples of Wearable Camera Images (Narrative Clip from NTCIR-13)

Typically, the datasets consist of:

- **Multimedia Content**: Wearable camera images captured at a rate of about two images per minute and worn from breakfast to sleep. Accompanying this image data for NTCIR-13/14 was a time-stamped record of music listening activities sourced from `Last.FM`[1] and (for NTCIR-14) an archive of all conventional (active-capture) digital photos taken by the lifelogger.
- **Biometrics Data**: Using off-the-shelf fitness trackers,[2] the lifeloggers gathered 24×7 heart rate, caloric burn and steps. In addition, for NTCIR-2014, continuous blood glucose monitoring was added which captured readings every 15 min using the Freestyle Libre wearable sensor.[3]
- **Human Activity Data**: The daily activities of the lifeloggers were captured in terms of the semantic locations visited, physical activities (e.g. walking, running, standing) from the Moves app,[4] along with (for NTCIR-14) a time-stamped diet log of all food and drink consumed.
- **Enhancements to the Data**: The wearable camera images were annotated with the outputs of various visual concept detectors which described in textual form the content of the lifelog images.

Readers who are interested in more information on the three lifelog datasets are referred to the task overview papers for NTCIR-12 (Gurrin et al. 2016), NTCIR-13 (Gurrin et al. 2017) and NTCIR-14 (Gurrin et al. 2019a). See Table 13.1 for a summary comparison of the three datasets.

What makes lifelog dataset generation a challenging task is the personal nature of real lifelog data (Chaudhari et al. 2007; Dang-Nguyen et al. 2017b) which must be gathered and released in a carefully organised process. One, or more, individuals must be willing to share a digital representation of their real-world activities with both researchers and the community. Aside from the difficulties of finding lifeloggers willing to share, various legal and institutional requirements needed to be met, such as passing review by an institutional ethics board, and for NTCIR-14, the preparation of a Data Protection Impact Assessment (to meet European GDPR requirements). Datasets were made available via the NTCIR-Lifelog website[5] and were password protected and secured by HTACCESS with username/password pairs generated for each participant. Additionally, in a style similar to TREC, each participating organisation needed an appropriate representative to sign an organisational agreement form and send it to the task organiser. Individual agreement forms were maintained by the participating organisation on behalf of each task participant within that organisation.

Prior to release, each dataset was subject to a detailed multi-phase redaction process to anonymise the dataset in terms of the lifelogger's identity as well as the identity of bystanders in the data. While many approaches have been proposed to

---

[1]Last.FM Music Tracker—https://www.last.fm/.

[2]For example, the Fitbit Fitness Tracker (FitBit Versa)—https://www.fitbit.com/.

[3]Freestyle Libre wearable glucose monitor—https://www.freestylelibre.ie/.

[4]Moves App for Android and iOS—http://www.moves-app.com/.

[5]NTCIR-Lifelog website—http://ntcir-lifelog.computing.dcu.ie/.

Table 13.1 Statistics of NTCIR lifelog datasets

Criteria	NTCIR-12	NTCIR-13	NTCIR-14
Number of Lifeloggers	3	2	2
Number of Days	90 days	90 days	43 days
Collection Size	18 GB	26 GB	14 GB
Number of Images	88,124 images	114,547 images	81,474 images
Number of Locations	130 locations	138 locations	61 locations
Physical Activities	Moves app	Moves app	Moves app
Calorie Burn	-	Fitness Watch	Fitness Watch
Step Count	-	Fitness Watch	Fitness Watch
Heart Rate	-	Chest Strap	Fitness Watch
Blood Glucose	-	Daily	Continuous
Music Listening	-	Last.FM	Last.FM
Cholesterol	-	Weekly	-
Uric Acid	-	Weekly	-
Diet Log	-	Manual	Manual
Conventional Photos	-	-	Smartphone

supporting privacy preservation in lifelog data (Gurrin et al. 2014a; Memon and Tanaka 2014), it was realised that none were effective enough to be deployed in an automated manner over lifelog data. Hence, a multi-step process was put in place that relied on manual (or semi-manual) redaction, and is summarised as follows:

- **Data Filtering:** Given the personal nature of lifelog data, it was necessary to allow the lifeloggers to remove any lifelog data that they may have been unwilling to share. This sharable data was then reviewed by a trusted member of the organising team and further deletions occurred where deemed prudent.
- **Privacy Protection:** Privacy-by-design (Cavoukian 2010) was a requirement for the test collection. Consequently, faces, readable screens and personal details (e.g. bank cards, passports) were blurred in either a fully manual or semi-automated process. Additionally, every image was resized down to 1024×768 resolution which had the effect of rendering most textual content illegible. Following this, a validation check was performed on the redaction outputs.

The overall data redaction and release process is summarised in Fig. 13.2, which shows the steps taken by the lifelogger (1), the organisers (2) and the responsibility on the task participants (3) who use the data for their experiments. As can be seen, the lifelogger gets the opportunity to review, filter and clean their data before the organisers carry out a secondary data review and cleaning, followed by the execution of a number of processes to ensure privacy of individuals associated with the dataset, followed by a final validation of the data before it is released for interested researchers who sign up to access the data.
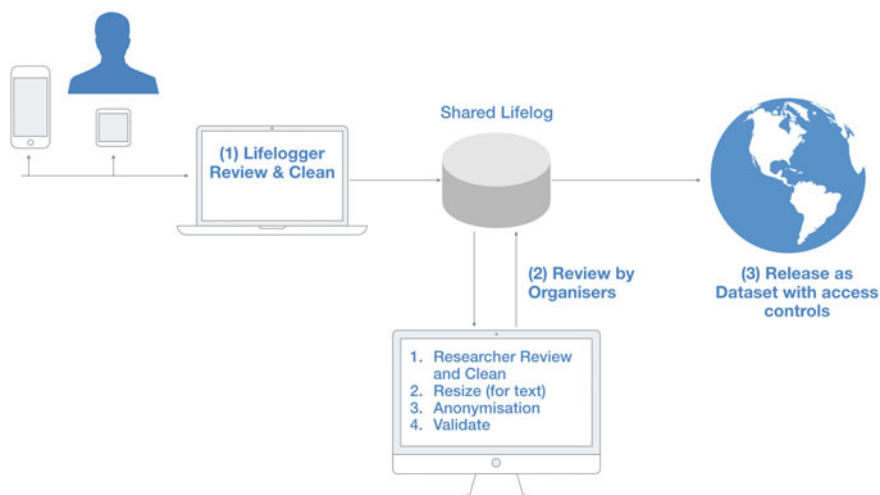
Fig. 13.2 Overview of the Redaction Process for the NTCIR Collections

13.4 Lifelog Subtasks at NTCIR

Based on the use cases described previously and guided by the human memory-access applications of Sellen and Whittaker (2010), five different challenges were explored at NRCIR-Lifelog. In this section, we focus on the two main subtasks that ran for all three Lifelog instances and we briefly describe the other three subtasks.

13.4.1 Lifelog Semantic Access Subtask

The Lifelog Semantic Access subtask (LSAT) was the core task of the three editions of the Lifelog task. The aim of the task was to explore ad hoc search and retrieval from lifelogs, which the organisers believe to be a fundamental enabling technology to make lifelogs a useful tool for individuals. In this subtask, the participants were required to retrieve a number of specific moments in a lifelogger's life in response to a topic description, as shown in Fig. 13.3. There were either 24 or 48 topics prepared for each instance of the task. For the purposes of evaluation, the organisers took the simplifying assumption that an image (point-in-time) is an appropriate document for retrieval. The task can best be compared to a known-item search task with one (or more) relevant items per topic. Evaluation was by means of standard evaluation measures and calculated using treceval.[6] For NTCIR-12 & NTCIR-13, full relevance judgements were prepared, but for NTCIR-14, pooled relevance judgements were

---

[6]https://trec.nist.gov/trec_eval/.

> TITLE: *Icecream by the Sea*
> DESCRIPTION: *Find the moment when U_1 was eating icecream beside the sea.*
> NARRATIVE: *To be relevant, the moment must show both the icecream with cone in the hand of u1 as well as the sea clearly visible. Any moments by the sea, or eating an icecream which do not occur together are not considered to be relevant.*
>
> *EXAMPLES OF RELEVANT MOMENTS FOUND BY PARTICIPANTS*
>
>

Fig. 13.3 LSAT Topic Example, including example results

used. Participants were allowed to undertake the LAST subtask in an interactive or automatic manner. For interactive submissions, a maximum of five minutes of search time was allowed per topic.

Over the three instances of the LSAT Task, we note that task participants took many different approaches to the development of retrieval systems. Given that there are no standardised baselines that can be applied, this is not surprising. Participating teams developed many different experimental systems, both interactive and automatic in nature. We look firstly at interactive retrieval engines over the three editions of NTCIR. At NTCIR-12, the participating team from University of Barcelona (Spain) developed the only interactive retrieval engine that integrated a semantic-content tagging tool to enhance the quality of the annotations (de Oliveira Barra et al. 2016). At NTCIR-13, the DCU team (Ireland) employed a human-in-the-loop to translate the provided queries into system queries for their retrieval engine, in one of their runs (Duane et al. 2017). However, at NTCIR-14, we note that three of the participants developed interactive systems and a fourth participant also integrated the human-in-the-loop query enhancement. NTU (Taiwan) developed an interactive lifelog retrieval system that automatically suggested to the user a list of candidate query words and adopted a probabilistic relevance-based ranking function for retrieval (Fu et al. 2019). They enhanced the official concept annotations and pre-processed the visual content to remove poor quality images and to offset the fish-eye nature of the wearable camera data. DCU (Ireland) developed an interactive retrieval engine for lifelog data (Ninh et al. 2019) that was designed for novice users and relied on an extensive list of facet filters over provided metadata. Finally, the VNU-HCM (Vietnam) group developed an interactive retrieval system (Nguyen et al. 2019) that used enhanced metadata

and visual enrichment, sometimes including human annotations. Their scalable and user-friendly interface to this system significantly outperformed competing systems at NTCIR-14, due primarily to the enhanced annotations. As expected, all interactive runs significantly outperformed the automatic runs at each edition of NTCIR-Lifelog.

In terms of approaches to automatic retrieval, at NTCIR-12, the VTIR (USA) team hypothesised that location was a very important component in the information retrieval process (Xia et al. 2016), and thus enhanced location semantic descriptions were used with the BM25 retrieval model. The authors comment that this approach worked well for some of the topics, which were location dependent. The IDEAS Institute for Information Industry (Taiwan) took a textual approach to retrieval (Lin et al. 2016) utilising word2vec to better match visual concepts to user queries (an approach referred to as bridging the lexical gap) via query expansion. The QUT group took an approach to retrieval that generated long, descriptive paragraphs of text to annotate the lifelog content, as opposed to the conventional tag-based approach (Scells et al. 2016); however, this was not shown to be successful. Finally, the LIG-MRM group (France) performed significantly better of all other approaches at NTCIR-12, by focusing on enhancing the performance of the visual concept detectors to be used for retrieval, and not relying on the provided (Caffe) classifier output (Safadi et al. 2016). The Caffe classifier provides a modifiable framework for state-of-the-art deep learning algorithms and a collection of reference models (Jia et al. 2014).

AT NTCIR-13, three participating groups took part in the LSAT subtask in an automated manner. DCU (Ireland) took part with their baseline search engine (Duane et al. 2017) that indexed the provided metadata and concepts using BM25 as the retrieval model, with both automated query runs and human-enhanced query runs. VCI2R (Singapore) proposed a general framework to bridge the semantic gap between lifelog data and the event-based LSAT topics (Lin et al. 2017) by enhancing the visual annotations and employing temporal smoothing of annotations, which proved to be the most successful approach at NTCIR-13. Finally, the PGB group (Japan) focused on the image and location data and enhanced the visual annotations (including people counting) and indexed locations using point-stay detection (D-Star algorithm) and integrated important location detection using the DBSCAN algorithm (Yamamoto et al. 2017). It performed better than the baseline, but not as well as the VCI2R and the human-in-the-loop run by DCU.

At NTCIR-14, NTU (Taiwan) submitted both interactive and automatic runs, and their automatic run (the top-ranked automatic run) included a query enhancement process using the top 10 nearest concepts to the query terms to expand the query before submitting the query (Fu et al. 2019). QUIK (Japan) from Kyushu University integrated online visual WWW content in the search process and operated based on an underlying assumption that a lifelog image of an activity would be similar to images returned from a WWW search engine for similar activities (Suzuki and Ikeda 2019). The approach operated using only the visual content of the collection and used the WWW data to train a visual classifier with a convolutional neural network for each topic. Although an automated process, a human-in-the-loop mechanism was employed to filter the WWW examples.

After NTCIR-14, the main approaches that the organisers consider to be valuable for lifelog access are the use of enhanced visual concept detectors to improve indexing, which has been continually shown to be effective both at NTCIR and the Lifelog Search Challenge (Gurrin et al. 2019b), as well as the application of approaches to bridging the lexical gap, either via some form of index term expansion or query expansion. Given the interest in developing interactive systems, the Lifelog Search Challenge is now the main venue for the comparative benchmarking of interactive lifelog retrieval systems.

13.4.2 Lifelog Insight Subtask

The Lifelog Insight subtask (LIT) also ran at all three editions of NTCIR-Lifelog and was designed to explore knowledge mining from lifelogs, with particular application in epidemiological studies. The LIT subtask 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 visualisation of knowledge mined from self-tracking data to provide 'self-knowledge through numbers'. Participants were requested to provide insights that support the lifelogger in the act of reflecting upon their life, facilitate filtering, or provide for efficient/effective means of lifelog data visualisation. The LIT subtask was not evaluated in the traditional sense, rather all participants were asked to write about and bring their demonstrations or reflective output at the NTCIR conference.

At NTCIR-12, the Sakai Lab at Waseda University (Japan) developed a prototype smartphone application called Sleepflower, which was designed to improve the sleep cycles of a group of users (Iijima and Sakai 2016). A flower metaphor was displayed on the smartphone screen to represent the current sleepiness of a particular user, based on a manual analysis of the habits of the lifeloggers. Participants from Toyohashi University (Japan) examined repeated pattern discovery from lifelog image sequences, by applying a Spoken Term Discovery technique (Yamauchi and Akiba 2016) and a variant of Dynamic Time Warping was used in an experimental approach to extract meaningful patterns from the lifelog data. DCU (Ireland) 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 (Duane et al. 2016). The results of this manual interrogation were then used to generate insights and infographics.

At NTCIR-13, Tsinghua University (China) developed an approach to give insights into the big-five personality traits, moods, music moods, style detection and sleep-quality prediction (Soleimaninejadian et al. 2017). The team augmented the provided dataset with lifelog data gathered by other volunteers. The team found that their approaches achieved objective results with a high degree of accuracy, and noted the implications for improving traditional psychological research by employing lifelog data. Participants from the Institute for Infocomm Research (Singapore) presented a method for finding insights from the lifelog data by creating a topic-

focused minute-by-minute annotation of the user's activities (Xu et al. 2017). This was achieved by applying deep learning approaches for image analytics and then fusing the multimodal sensor data to generate insights into patterns and associations between lifelogger activities. The team from DCU (Ireland) introduced a new interactive lifelog interrogation system which was implemented for access in a Virtual Reality Environment (Duane et al. 2017). The system was designed to allow a user to explore visual lifelog data in an interactive and highly visual manner. Finally, the PGB group (Japan) developed an approach to automatically label the lifelog images with 15 concept labels (Yamamoto et al. 2017) using a DNN model with a fusion layer of tri-modal data (image, location and biometric).

At NTCIR-14, only one group took part in the LIT subtask. THUIR (China) developed a number of detectors for the lifelog data to automatically identify and visualise the status/context of a user (Nguyen et al. 2019) and a comparison between the various approaches showed that the visual features were significantly better than non-visual (metadata) features.

13.4.3 Other Subtasks (LEST, LAT and LADT)

A number of additional exploratory subtasks were run once (or twice) only. We will briefly describe these and comment on why they were not run in all three instances of the Lifelog task. The Lifelog Event Segmentation subtask (LEST) ran at NTCIR-13, the aim of which was to examine approaches to event segmentation from continual lifelog stream data (Gurrin et al. 2017). Event segmentation had been the typical approach to generation of indexable and retrievable documents (events) from lifelog collections. Given that the definition of an event is inherently subjective to the experience of the individual lifelogger, the organisers defined 15 types of events for the segmentation process, based on the 15 common lifestyle activities defined by Kahneman et al. (2004). The PGB group (NTT, Japan) participated in the LEST and developed a number of alternative approaches to event segmentation, included temporal visual similarity, user-linger-points, the use of LDA to reduce dimensionality and identify boundaries, and a multi-feature approach that used cosine similarity between segments (Yamamoto et al. 2017). The user-linger-points approach proved to be the most successful for event segmentation.

At NTCIR-14, this LEST morphed into the Lifelog Activity Detection subtask (LADT) at NTCIR-14 (Gurrin et al. 2019a), which required the classification of the multimodal lifelog data into one or more human activities that were identified as occurring in the lifelog collection. The NTU group (Taiwan) developed a new approach for the multi-label classification of lifelog images (Fu et al. 2019). In order to train the classifier, the authors manually labelled 4 days, which were chosen because they covered most of the activities that the lifeloggers were involved in.

However, the organisers note that there was little interest from the community in this task. This was surprising, since many of the previous applications of lifelog data to solve real-world challenges (e.g. healthcare or epidemiological studies) would

require the detection of human activities as a fundamental building block. Perhaps, this task will become very relevant and interesting at a later date, once lifelogging becomes a more commonplace activity for personal use or scientific enquiry.

It is worth noting that one outcome of this subtask was a new pilot task at NTCIR-15, which has a micro-activity detection/retrieval task (called MART) that extends this early work by focusing on the identification of short activities of daily life (e.g. writing an email, making a cup of coffee) and is targeted at the generation of rich and detailed semantic logs of everyday activities.

Finally, another exploratory subtask that ran at NTCIR-13 was the Lifelog Annotation subtask (LAT), which aimed to develop approaches for annotation of the multimodal lifelog data (images) with a fixed set of 15 high-level labels/concepts chosen from a manually generated ontology of lifelogging activities (Gurrin et al. 2017). These concepts were based on both the activities (facets of daily life) of the individual and the environmental settings (contexts) of the individual. Motivated by the realisation from NTCIR-12 that high-quality annotations are important for the retrieval process, the aim of this task was to provide various sets of high-quality shared annotations for all other uses to use in the LSAT subtask. However, only one group participated, so this annotation sharing did not occur. The PGB group (Yamamoto et al. 2017) developed a DNN model, with a fusion layer of tri-modal data (image, location and biometrics) to perform the content annotation. It was found that visual and biometric features can enhance the automatic annotation process, yet location actually was found to reduce annotation quality. Once again, this task was not attractive to NTCIR participants, so the Lifelog Activity Detection subtask (LADT) at NTCIR-14 replaced it.

13.5 Lessons Learned

Since NTCIR-12, 18 different research groups have taken part in the Lifelog task, some of them multiple times and across multiple tasks. Uptake on the subtasks suggests that the community is interested in the retrieval challenge and, to a lesser extent, the insights challenge. The other three challenges have not attracted much interest at this point. At the end of the NTCIR-Lifelog tasks, we can identify some lessons learned from the three editions of the NTCIR-Lifelog task:

- Novel Datasets: Eighteen participants submitted official runs to NTCIR, but at least three times as many downloaded the datasets. Even 4 years after starting the NTCIR-Lifelog task, requests for the datasets are still being received by the organisers. There is clearly an interest in the community to develop retrieval and analytics tools over such datasets, so there is significant potential for others in the community to define and release novel datasets of human life-experience data.
- Richer metadata: Repeatedly, we have seen that the best performing retrieval systems enhanced the provided metadata by relying on additional visual concept detectors, or seeking additional sources of metadata to enhance the retrieval performance. There is clearly a need to develop new approaches to the creation of

semantically rich metadata for multimodal lifelogs, in order to facilitate more effective retrieval algorithms.

- Bridge the lexical gap: Many participants found that there was a lexical gap between the terms used by the lifeloggers in their topic descriptions, and the indexed textual content and annotations. This suggests a need for term or query expansion, and the current consideration is that this could be achieved using approaches such as conventional query expansion or word embedding.
- Integrate external WWW content: This has been used by some participants with positive results. The external content helps to enhance the quality of content annotations or can be used as a form of query enhancement.
- There is an observed interest in the generation of insights or knowledge from lifelog data, as seen by the participation in the LIT subtask. This seems best suited to addressing the reflection and reminiscence use case of human memory as outlined by Sellen and Whittaker (2010).
- Document segmentation of the lifelog data into indexable content is as of yet an unsolved challenge. Initial attempts at lifelog 'event segmentation' (Lee et al. 2008) generated static documents for retrieval using an early sensor-based approach to segmentation. As with any information retrieval system, the concept of a document needs to be clearly defined and understood, which is not yet the case for lifelog data.
- Interactive search: Finally, interactive systems have been increasing in interest since NTCIR-12 and the Lifelog Search Challenge (Gurrin et al. 2019b) has been started to specifically explore this challenge. This appears to be the current hot topic for lifelog search and retrieval.

13.5.1 Conclusions and Future Plans

Over the course of the three instances of the NTCIR-Lifelog task, the uptake by participants was not as high as the organisers had hoped. One reason for this may be the emergence of a suite of parallel activities to motivate research into lifelogging and personal data analytics, such as the previously introduced interactive Lifelog Search Challenge (Gurrin et al. 2019b) and the ImageCLEF-Lifelog activities (Ionescu et al. 2018). The Lifelog Search Challenge in particular his been attracting 8–10 groups annually who come together to partake in a real-time interactive search challenge, which provides an open forum for all ACM ICMR conference attendees to partake as either observers or even as novice users in the competition. The ImageCLEF-Lifelog task tends to attract researchers more focused on the computer vision aspects of insight generation and data organisation and as such, it is targeting a slightly different audience. Regardless of the reasons, the uptake of the task and the level of interest in the dataset, along with the other related activities suggests a keen level of interest in the community for lifelog retrieval and the organisers note that this interest is likely to grow as volumes of personal multimodal data increase in society. The organisers understand that lifelog retrieval is a challenging activity,

and the future of the Lifelog task at NTCIR is perhaps in the refinement of the task to address key challenges in the domain, such as privacy-aware retrieval from personal multimodal data, epidemiological-scale analytics studies that analyse large lifelogs from multiple participants, targeted healthcare tasks of interest to concerned individuals and medical professions (e.g. finding medicine-taking events), or novel related-domains such as neural data retrieval.

It is an inevitable fact that the main challenge for any organisers of such tasks is the effort required to generate appropriate and real-world datasets and release them in an ethically and legally complaint manner. The three lifelog datasets released by the task organisers at NTCIR represent about a year of effort in total from a number of researchers and lifeloggers; this naturally incurs significant expenses in terms of organisers time and resources. Real-world use cases are likely to either focus on retrieval from longitudinal archives donated by one individual, or across large populations (as in epidemiological studies) and the data gathering and release methodology employed for this task was not ideal, due to the large overhead of effort required to ensure privacy preservation. The evaluation-as-a-service model proposed by Hopfgartner et al. (Hopfgartner et al. 2020) is one potential way forward, which brings the algorithms to the data, rather than the conventional data-to-algorithm approach. Another potential next step is to encourage more comparative evaluation of interactive systems, since a user of a lifelog tool (either an individual or a professional analyst) is most likely to be using such tools in an interactive manner. In any case, the organisers of the NTCIR-Lifelog tasks consider that this book chapter marks the end-of-the-beginning of research into lifelog data organisation and retrieval, rather than the conclusion of a short-lived sub-topic of IR. It is our belief that lifelogging as a topic will continue to become more popular for IR researchers and that the availability of relevant datasets and challenges will increase in the coming years.

References

Barnard PJ, Murphy FC, Carthery-Goulart MT, Ramponi C, Clare L (2011) Exploring the basis and boundary conditions of SenseCam-facilitated recollection. Memory 19(7):758–767

Berry E, Kapur N, Williams L, Hodges S, Watson P, Smyth G, Srinivasan J, Smith R, Wilson B, Wood K (2007) The use of a wearable camera, SenseCam, as a pictorial diary to improve autobiographical memory in a patient with limbic encephalitis: a preliminary report. Neuropsychol Rehabil 17(4–5):582–601

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

Chaudhari J, Cheung SS, Venkatesh M (2007) Privacy protection for life-log video. In: IEEE workshop on signal processing applications for public security and forensics, 2007 SAFE '07, pp 1–5

Chowdhury S, Ferdous MS, Jose JM (2016) A user-study examining visualization of lifelogs. In: 2016 14th International Workshop on Content-Based Multimedia Indexing (CBMI), pp 1–6

Dang-Nguyen DT, Piras L, Riegler M, Boato G, Zhou L, Gurrin C (2017a) Overview of ImageCLE-Flifelog 2017: lifelog retrieval and summarization. In: CLEF2017 working notes, CEUR-WS.org, Dublin, Ireland, CEUR workshop proceedings

Dang-Nguyen DT, Zhou L, Gupta R, Riegler M, Gurrin C (2017b) Building a disclosed lifelog dataset: Challenges, principles and processes. In: Proceedings of the 15th international workshop on content-based multimedia indexing. ACM, New York, NY, USA, CBMI '17, pp 22:1–22:6

Dang-Nguyen DT, Piras L, Riegler M, Zhou L, Lux M, Gurrin C (2018) Overview of ImageCLE-Flifelog 2018: daily living understanding and lifelog moment retrieval. In: CLEF2018 working notes. CEUR workshop proceedings, Avignon, France. Accessed 10–14 Sept 2018

Dodge M, Kitchin R (2007) 'Outlines of a world coming into existence': pervasive computing and the ethics of forgetting. Environ Plann B: Plann Des 34(3):431–445

Doherty AR, Moulin CJa, Smeaton AF, (2011) Automatically assisting human memory: A Sense-Cam browser. Memory 19(7):785–795

Doherty AR, Pauly-Takacs K, Caprani N, Gurrin C, Moulin CJa, O'Connor NE, Smeaton AF (2012) Experiences of aiding autobiographical memory using the sensecam. Human-Comput Interac 27(1–2):151–174

Duane A, Gurrin C, Zhou L, Gupta R (2016) Visual insights from personal lifelogs - insight at the NTCIR-12 lifelog LIT task. In: Proceedings of NTCIR-12, Tokyo, Japan

Duane A, Zhou L, Dang-Nguyen DT, Gurrin C (2017) DCU at the NTCIR-13 lifelog-2 task. In: Proceedings of NTCIR-13, Tokyo, Japan

Everson B, Mackintosh KA, McNarry MA, Todd C, Stratton G (2019) Can Wearable Cameras be Used to Validate School-Aged Children's Lifestyle Behaviours? Children 6(2):20

Fu MH, Chia-Chun C, Huang GH, Chen HH (2019) Introducing external textual knowledge for lifelog retrieval and annotation. In: Proceedings of NTCIR-14, Tokyo, Japan

Gemmell J, Bell G, Lueder R, Drucker S, Wong C (2002) Mylifebits: Fulfilling the memex vision. In: Proceedings of the Tenth ACM International Conference on Multimedia, ACM, New York, NY, USA, ACM Multimedia '02, pp 235–238

Gurrin C, Albatal R, Joho H, Ishii K (2014a) A privacy by design approach to lifelogging. In: O Hara K, Nguyen C, Haynes P (eds) Digital enlightenment yearbook 2014. IOS Press, pp 49–73

Gurrin C, Smeaton AF, Doherty AR (2014b) LifeLogging: personal big data. Found Trends® Inf Retriev 8(1):1–125

Gurrin C, Joho H, Hopfgartner F, Zhou L, Albatal R (2016) Overview of NTCIR-12 lifelog task. In: Proceedings of NTCIR-12, Tokyo, Japan

Gurrin C, Joho H, Hopfgartner F, Zhou L, Gupta R, Albatal R, Dang-Nguyen DT (2017) Overview of NTCIR-13 lifelog-2 task. In: Proceedings of NTCIR-13, Tokyo, Japan, pp 6–11

Gurrin C, Joho H, Hopfgartner F, Zhou L, Ninh VT, Le TK, Albatal R, Dang-Nguyen DT, Healy G (2019a) Overview of NTCIR-14 lifelog-3 task. In: Proceedings of NTCIR-14, Tokyo, Japan

Gurrin C, Schoeffmann K, Joho H, Zhou L, Duane A, Leibetseder A, Riegler M, Piras L (2019b) Comparing approaches to interactive lifelog search at the lifelog search challenge (LSC2018). ITE Trans Media Technol Appl 7(2):46–59

Harvey M, Langheinrich M, Ward G (2016) Remembering through lifelogging: a survey of human memory augmentation. Pervasive Mobile Comput 27:14–26

Hopfgartner F, Gurrin C, Joho H (2020) Rethinking the test collection methodology for personal self-tracking data. In: 26th international conference on multimedia modeling 2020. Springer International Publishing

Hoyle R, Templeman R, Armes S, Anthony D, Crandall D, Kapadia A (2014) Privacy behaviors of lifeloggers using wearable cameras. In: Proceedings of the 2014 ACM international joint conference on pervasive and ubiquitous computing - UbiComp '14 Adjunct, pp 571–582

Iijima S, Sakai T (2016) SLLL at the NTCIR-12 lifelog task: sleepflower and the LIT subtask. In: Proceedings of NTCIR-12, Tokyo, Japan

Ionescu B, Müller H, Villegas M, de Herrera AGS, Eickhoff C, Andrearczyk V, Cid YD, Liauchuk V, Kovalev V, Hasan SA et al (2018) Overview of ImageCLEF 2018: Challenges, datasets and evaluation. International Conference of the Cross-Language Evaluation Forum for European Languages. Springer, Cham, pp 309–334

Jia Y, Shelhamer E, Donahue J, Karayev S, Long J, Girshick R, Guadarrama S, Darrell T (2014) Caffe: convolutional architecture for fast feature embedding. arXiv:1408.5093

Kahneman D, Krueger AAB, Schkade DA, Schwarz N, Stone AAA (2004) A survey method for characterizing daily life experience: the day reconstruction method. Science 306(5702):1776–1780

Lee H, Smeaton AF, O'Connor NE, Jones G, Blighe M, Byrne D, Doherty A, Gurrin C (2008) Constructing a SenseCam visual diary as a media process. Multimed Syst 14(6):341–349

Lin HL, Chiang TC, Chen LP, Yang PC (2016) Image searching by events with deep learning for NTCIR-12 lifelog. In: Proceedings of NTCIR-12, Tokyo, Japan

Lin J, del Molino AG, Xu Q, Fang F, Subbaraju V, Lim JH (2017) VCI2R at the NTCIR-13 lifelog semantic access task. In: Proceedings of NTCIR-13, Tokyo, Japan

Lokoc J, Bailer W, Schoeffmann K, Münzer B, Awad G (2018) On influential trends in interactive video retrieval: video browser showdown 2015–2017. IEEE Trans Multimed 20(12):3361–3376

Memon MA, Tanaka J (2014) Ensuring privacy during pervasive logging by a passerby. J Inf Process 22(2):334–343

Milton F, Muhlert N, Butler CR, Smith A, Benattayallah A, Zeman A (2011) An fMRI study of long-term everyday memory using SenseCam. Memory 19(7):733–744

Nguyen IVK, Shrestha P, Zhang M, Liu Y, Ma S (2019) THUIR at the NTCIR-14 lifelog-3 task: how does lifelog help the users status recognition. In: Proceedings of NTCIR-14, Tokyo, Japan

Nguyen THC, Nebel JC, Florez-Revuelta F (2016) Recognition of activities of daily living with egocentric vision: a review. Sensors 16(1):

Ninh VT, Le TK, Zhou L, Healy G, Venkataraman K, Tran MT, Dang-Nguyen DT, Smith S, Gurrin C (2019) A baseline interactive retrieval engine for the NTICR-14 Lifelog-3 semantic access task. In: Proceedings of NTCIR-14, Tokyo, Japan

de Oliveira Barra G, Ayala AC, Bolaos M, Dimiccoli M, Aghaei M, Carn M, Giro-I-Nieto X, Radeva P (2016) LEMoRe: a lifelog engine for moments retrieval at the NTCIR-lifelog LSAT task. In: Proceedings of NTCIR-12, Tokyo, Japan

Safadi B, Mulhem P, Quénot G, Chevallet JP (2016) MRIM-LIG at NTCIR lifelog semantic access task. In: Proceedings of NTCIR-12, Tokyo, Japan

Scells H, Zuccon G, Kitto K (2016) QUT at the NTCIR lifelog semantic access task. In: Proceedings of NTCIR-12, Tokyo, Japan

Sellen AJ, Whittaker S (2010) Beyond total capture: a constructive critique of lifelogging. Commun ACM 53(5):70–77

Signal LN, Stanley J, Smith M, Barr MB, Chambers TJ, Zhou J, Duane A, Gurrin C, Smeaton AF, McKerchar C, Pearson AL, Hoek J, Jenkin GLS, Ni Mhurchu C (2017) Children's everyday exposure to food marketing: an objective analysis using wearable cameras. Int J Behav Nutr Phys Activity 14(1):137

Soleimaninejadian P, Wang Y, Tong H, Feng Z, Zhang M, Liu Y, Ma S (2017) Bridging technology and psychology through the lifelog. In: Proceedings of NTCIR-13, Tokyo, Japan

Suzuki T, Ikeda D (2019) Smart lifelog retrieval system with habit-based concepts and moment visualization. In: Proceedings of NTCIR-14, Tokyo, Japan

Wilson G, Jones D, Schofield P, Martin DJ (2018) The use of a wearable camera to explore daily functioning of older adults living with persistent pain: methodological reflections and recommendations. J Rehabil Assist Technol Eng 5(205566831876):541

Xia L, Ma Y, Fan W (2016) VTIR at the NTCIR-12 2016 lifelog semantic access task. In: Proceedings of NTCIR-12, Tokyo, Japan

Xu Q, Subbaraju V, del Molino AG, Lin J, Fang F, Lim JH (2017) Visualizing personal lifelog data for deeper insights at the NTCIR-13 lifelog task. In: Proceedings of NTCIR-13, Tokyo, Japan

Yamamoto S, Nishimura T, Akagi Y, Takimoto Y, Inoue T, Toda H (2017) PBG at the NTCIR-13 lifelog-2 LAT, LSAT, and LEST tasks. In: Proceedings of NTCIR-13, Tokyo, Japan

Yamauchi K, Akiba T (2016) 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

Zhou Q, Wang D, Mhurchu CN, Gurrin C, Zhou J, Cheng Y, Wang H (2019) The use of wearable cameras in assessing children's dietary intake and behaviours in China. Appetite. https://doi.org/10.1016/J.APPET.2019.03.032