



VRLE: Lifelog Interaction Prototype in Virtual Reality

Lifelog Search Challenge at ACM ICMR 2020

Aaron Duane
 Dublin City University
 Dublin, Ireland
 aaron.duane@dcu.ie

Björn Þór Jónsson
 IT University of Copenhagen
 Copenhagen, Denmark
 bjth@itu.dk

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

ABSTRACT

The Lifelog Search Challenge (LSC) invites researchers to share their prototypes for interactive lifelog retrieval and encourages competition to develop and evaluate effective methodologies to achieve this. With this paper we present a novel approach to visual lifelog exploration based on our research to date utilising virtual reality as a medium for interactive information retrieval. The VRLE prototype presented is an iteration on a previous system which won the first LSC competition at ACM ICMR 2018.

KEYWORDS

Lifelog, interactive retrieval, virtual reality.

ACM Reference Format:

Aaron Duane, Björn Þór Jónsson, and Cathal Gurrin. 2020. VRLE: Lifelog Interaction Prototype in Virtual Reality: Lifelog Search Challenge at ACM ICMR 2020. In *Proceedings of the Third Annual Workshop on the Lifelog Search Challenge (LSC '20), June 9, 2020, Dublin, Ireland*. ACM, New York, NY, USA, 6 pages. <https://doi.org/10.1145/3379172.3391716>

1 INTRODUCTION

In this paper we introduce an iteration of the Virtual Reality Lifelog Explorer (VRLE) [8] which participated in and won the first Lifelog Search Challenge (LSC) held at ACM ICMR in 2018. Unlike the majority of lifelog applications, which typically appear on personal computers, and a few other mainstream platforms [18, 21], VRLE is an application designed for lifelog interaction in a virtual environment, specifically on the HTC Vive.

Despite the relative proliferation of virtual reality hardware in recent years, VR applications, especially outside of the entertainment industry, remain in their infancy. The primary aim of designing and developing the original VRLE prototype was to encourage viable alternative hardware platforms within the lifelog community. As such, our current VRLE system remains a novel first-generation VR prototype that translates common lifelog interaction techniques [12] from more conventional media rather than establishing an entirely novel interaction paradigm.

With the successful deployment of the VRLE prototype as a proof of concept, our future work is intended to produce a next-generation virtual reality lifelog explorer based on the knowledge gained. To assist in this development, we have modified our original prototype with new features and interactions which we intend to evaluate via the third Lifelog Search Challenge at ACM ICMR 2020. In this paper we describe these additions by first discussing the background of virtual reality development before moving on to the design of our 3D user interface and how it supports lifelog query generation in a virtual space. This will be followed by a discussion on our more recently developed features, namely event ranking, visualisation, keyframe selection and the underlying search engine which supports these processes.

2 BACKGROUND

Our research to date has focused on examining the potential of virtual reality to support intuitive interaction with lifelog archives and we have developed several variations of the VRLE prototype to evaluate the feasibility of exploring lifelogs in virtual environments [7]. This work is motivated by our belief that virtual reality platforms will become more lightweight and ubiquitous in the near-future.

Effectively addressing the retrieval of lifelog documents (images, tweets, emails, etc.) has seen much attention in the lifelogging community in the last decade. This can be clearly observed in participation conferences such as the NTCIR Lifelog task [10] where an international cohort of researchers independently develop tools to effectively retrieve lifelog documents in media-rich personal datasets released by the challenge organisers. Some of this work has focused on enhancing the performance of the visual concept detectors to be used for retrieval [19] whereas other work has focused on a purely textual approach [15]. Other notable research has utilised long, descriptive paragraphs of text to annotate the lifelog content, as opposed to the conventional automatic tag-based approach [20] and some even created interactive systems which employed a semantic content-tagging mechanism [2]. Event segmentation [4] has also been a popular method of enhancing the retrieval process. Aside from the VRLE prototype, no submissions to an NTCIR conference have considered virtual reality as a platform for these methods of lifelog retrieval.

Though there has been almost no other research explicitly targeting the exploration of lifelogs in virtual reality, there have been many applications developed for the platform that facilitate elements of exploring and examining life experiences. One obvious example is the playback of 360-degree video which is considerably more immersive when viewed in virtual reality and is especially so when the footage is recorded from a more familiar first-person

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

LSC '20, June 9, 2020, Dublin, Ireland

© 2020 Association for Computing Machinery.

ACM ISBN 978-1-4503-7136-0/20/06.

<https://doi.org/10.1145/3379172.3391716>



Figure 1: Concept Filter



Figure 2: Day/Hour Filter

perspective. This evolution of immersion within virtual reality extends to many interaction methodologies that would better facilitate lifelog exploration. This is not to suggest that explicit examples of lifelog interaction in virtual reality do not already exist. For example, an art installation by Alan Kwan titled ‘Bad Trip’ [13] was developed in 2012 which enables users to explore a manifestation of the creator’s mind and life experience within virtual reality.

3 USER INTERFACE

The VRLE virtual user interface maintains the same design and layout as our previous prototype, which remains unchanged due to it already utilising knowledge gained from previous interactive evaluations [5], and is outside the scope of research being conducted for this challenge. The virtual interface consists of two primary components which, for the sake of completeness, we will briefly re-describe in this section of the paper. The querying component was a virtual 3D interface designed to provide a quick and efficient means for a user to generate a filter query within the virtual environment. While there are many approaches that one could take to input queries, the focus in this work was on gesture-based interaction, as opposed to other forms of interaction such as voice-based.

The gesture-based querying interface consists of two sub-menus, one for selecting lifelog concepts of interest and the second for selecting the temporal aspect of the query (e.g. hours of the day or days of the week). A typical query to the system, such as ‘using the computer on a Saturday afternoon’ would require the user to use the concept sub-menu to select the appropriate visual descriptors (e.g. computer or laptop) and the temporal sub-menu to select the time range (afternoon) and the day of the week (Saturday). The user then hits the submit button, the query is executed and the result is displayed for the user to browse.

The concept sub-menu is shown in Figure 1 and the temporal sub-menu is shown in Figure 2. The querying interface is available for the user to access at any time by pressing a dedicated button on either of the two VR handsets. When the user submits their query, the interface disappears and the user is free to explore or browse the results within the virtual space. The concepts were divided into pages corresponding to their first letter and organised alphabetically on each page from left to right and top to bottom. The user can select no concepts or anywhere up to ten concepts

per filter query. The temporal sub-menu presents the user with the 7 days of the week and the 24 hours of the day. These days and hours can be selected in any combination to generate a temporal filter on the search results.

A notable aspect of the development of the original VRLE prototype was to establish an effective mode of interaction with a 3D user interface within the virtual environment. The primary goal of the VRLE was to retrieve specific images captured by the lifelogger quickly and efficiently and we evaluated different interaction paradigms within this context. The relative infancy of virtual reality as an interactive platform means that there is no definitive answer on how to best interact with a 3D interface in this context. There are no well-defined or understood interaction best practices to integrate (e.g. point-and-click in the desktop environment, or sweep-a-finger in a touchscreen environment). Our research to date has culminated in two primary methods of user interaction which we refer to as distance-based and contact-based virtual reality interface paradigms. We have included both interaction methods as options for this iteration of the VRLE prototype.

3.1 Distance-Based Interaction

The distance-based interface paradigm utilises interactive beams which originate at the top of the user’s wireless controllers. These beams are projected when the controllers are pointed at any relevant interface in the virtual environment and directly interact with that interface’s elements (see Figure 3). This method of interaction is comparable to a lean-back style of lifelog browsing [11] and is functionally similar to using a television remote or other such device. Pressing a button on the controller selects the concept or time-range that is being pointed at. Naturally, it is possible to use both hands to select concepts in parallel, should a sufficiently dexterous user be generating queries.

3.2 Contact-Based Interaction

The contact-based interface paradigm utilises a much more direct form of interaction, where the user must physically touch the interface elements with their controllers. To facilitate this process, the controllers are outfitted with a virtual drumstick-like appendage protruding from the head of each controller (see Figure 4). This object was added to enhance precision and fidelity when contacting

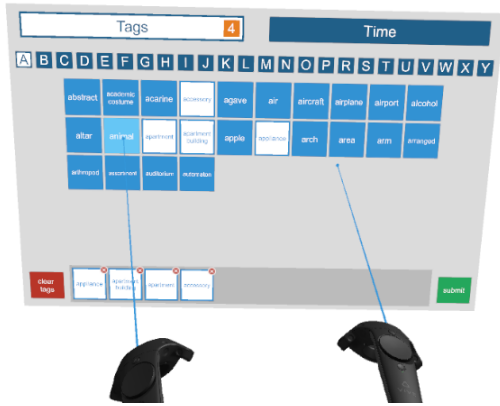


Figure 3: Distance-Based User Interaction

interface elements. This method of interaction is reminiscent of a more conventional style of lifelog browsing where the controller drumstick mimics how our fingers interact with a keyboard or touchscreen. Tactile feedback is provided through the handsets to signify hitting the keys.

3.3 Summary

The virtual reality interaction paradigms described in the preceding section are based on real-world analogues (television, keyboard, touchscreen, etc.) and can be observed in various forms within industry-standard virtual reality applications such as the HTC Vive's main menu¹ or Google's popular Tilt Brush palette.² To date we have not observed a clear user preference for a specific interface paradigm and there has been no notable advantage to using one interaction method over another in the context of lifelog retrieval [6].

4 DATA VISUALISATION

Similar to the user interface, the data visualisation approach in this iteration of the VRLE prototype shares many similarities with our original prototype from 2018. The primary difference is a focus on event-based visualisation around visual similarity of the lifelog images. The precise method of event generation is provided by and described in Duyen et al. [9]. Like the majority of lifelog datasets aiming for total capture, the primary focus of the LSC test collection is on continuous streams of images captured from the perspective of a lifelogger. Any one of these thousands of images could represent a potential cue to promote autobiographical memory, or in the context of the LSC, serve as the goal of a known-item search task.

For this latest iteration of the VRLE prototype we are experimenting with an event-based visualisation strategy incorporating the concept and temporal data as we believe it could serve as an effective strategy to support lifelog retrieval in virtual reality. This decision was influenced by the fact that event-based visualisation strategies have been implemented on a number of conventional lifelog applications in the past to varying success [4, 14], but their

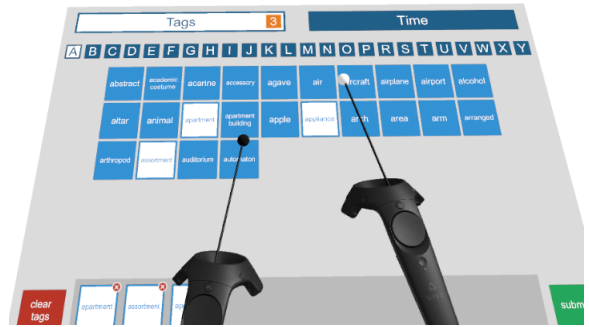


Figure 4: Contact-Based User Interaction

effectiveness in a virtual environment has yet to be fully explored. It is our hypothesis that the increase in accessible dimensions, highly immersive quality, and ability to coexist in the same space as the visualised data, will have a positive impact on the presentation of events and that this will in turn positively impact lifelog retrieval.

4.1 Ranking Events

An event is a sequence of lifelog images and metadata that corresponds to semantically and temporally related real-life activities. The search engine which supports the ranking of events in this iteration of the VRLE prototype is provided by Duyen et al. [9], which will also compete in LSC 2020 with a different system. The search engine follows an architecture which is typical of state-of-the-art lifelog search engines [22] as depicted in Figure 5. Initially, crucial features are extracted from images in the target dataset before they are formatted and indexed within an Elasticsearch engine. Elasticsearch is an open-source search and analytics engine built on Apache Lucene, which is a full-text search library. However, Elasticsearch can also index and retrieve large-scale data in diverse formats, including geospatial data, dates, numeric vectors, and images. All data, both structured and non-structured, needs to be transformed to JSON format in which the keys are the names of properties with their corresponding values to be stored in the engine. Duyen et al. utilise these advantages of Elasticsearch to index all necessary visual concepts, times, and GPS coordinates of

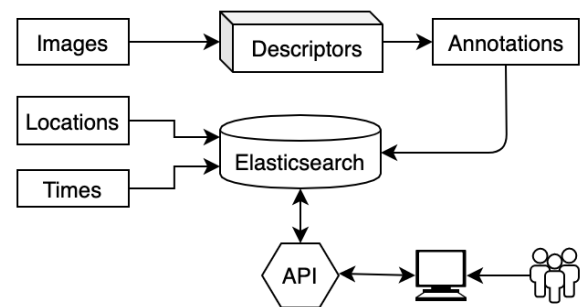


Figure 5: Search engine pipeline [9].

¹HTC Vive SteamVR, see: <http://store.steampowered.com/steamvr>.

²<https://www.tiltbrush.com/>



Figure 6: Ranked list of Event Previews

images as text, dates, and geospatial format, accordingly. In this version of their search engine, they have not used additionally released data such as music or heart rate, as this information did not notably contribute to previous LSC challenges. For more detailed information regarding this search engine please refer to the Inquirer prototype [9].

4.2 Visualising Events

Now that we have established a method of ranking events, we can begin to discuss a method to visualise these events. However, since the goal of an LSC task is not an event itself, but rather one of the images it contains, it is necessary to somehow expose these images to the user. Yet due to the quantity of visually similar images within the average event, displaying every image to the user at one time has the potential to inhibit effective retrieval rather than assist it. A simple solution to this is to summarise an event's content using a subset of its images which can then be further explored if the user considers the event to be relevant. In this instance we must also determine which images best convey the content and potential relevancy of an event, and how many should be exposed before it becomes inhibiting.

Through informal testing, it quickly became apparent that above a certain threshold, increasing the number of images in an event's summary only served to increase the time necessary to visually digest it without notable benefits to discerning its potential relevancy. This threshold was determined to be as few as three or four images for some events whereas others benefited from as many as six or seven. Continuing to increase this number resulted in a rapid decrease in efficiency, where users spent more time having to examine events or even began to ignore images completely. Though there is potential for future work in this area, it was decided that a nine image summary would adequately convey the relevancy of the majority of events in the dataset without notably impacting efficiency and also could be conveniently arranged in a 3x3 grid for ease of viewing (see Figure 6).

The design of this horizontal grid arrangement was based on a number of criteria. First, unlike conventional applications, the data needed to scroll horizontally rather than vertically. This is because positioning virtual elements below a user can convey a



Figure 7: Summary of one event with ranked keyframes

sense of being high off the ground which causes severe vertigo for some individuals. Second, the size of the virtual images needed to be large enough to retain their clarity when positioned within the user's reference frame but small enough so as not to dominate the user's visual range. Third, once an appropriate size of image was determined, we needed to establish the maximum number which could be displayed vertically without demanding excessive head movement from the user. We determined this number to be three, as using fewer than three images poorly utilised the available virtual space and using more than three images demanded notably more head movement.

4.3 Event Keyframe Selection

Now that we have established an appropriate size for the event summaries, we must determine a method of selecting appropriate images to effectively convey the event's content and potential relevancy. We refer to the images in our summary as event keyframes and, though there have been several approaches to keyframe selection within video retrieval [1, 17], there has been little work on keyframe selection within the lifelogging domain. For example, [3, 14] describe a selection of keyframe selection techniques, but their scope is restricted to extracting a single keyframe to represent an event.

We refer to our method of selecting keyframe images as the ranked event summary. This is because, although the images in the summary are arranged temporally from earliest to latest, the keyframe selection prioritises images with concept labels related to the user's original query. The goal was to provide a summary of the event with a bias toward images containing concept descriptors the user has queried. Specifically, this approach compiles a list of every image in the event ranked by the number of the queried concepts each image contains. If there are nine or more images in this ranked list, the top nine images are chosen for the summary of the event. However, when there are fewer than nine images, the system selects keyframe images equally spaced around the ranked images which have already been selected (see Figure 7).



Figure 8: Exploring Event with Image Metadata

4.4 Event Contextual Interfaces

With the introduction of the event preview summary, we now require a mechanism by which users can explore an event in its entirety. To achieve this, along with supporting other possible feature implementations, we established the concept of a contextual interface that would only be exposed when necessary. This type of contextual interface is common in event-based lifelog applications [4, 14] where it is typically achieved simply by hovering over relevant images. An initial design overlaid the contextual interface elements on target images within the virtual space, but after informal testing this was determined to be unsuitable as it became clear the overlaid elements were obscuring the data the user was trying to examine and negatively impacting retrieval.

One solution to this was to redesign the interface elements to be smaller and less obtrusive, but small icons and text rendered within the virtual environment needed to be within a specific range to remain clear and legible to users because the resolution afforded by the virtual reality hardware is low enough that aliasing (jagged edges) can occur after a short distance. Since users were capable of moving around the virtual space and adjusting their position with respect to the data, we could not rely on them always being within the appropriate range to render interface elements clearly.

Instead, the contextual interface was redesigned so that it would appear alongside the user's active controller, which meant that it would always appear within the appropriate range. In Figure 8 we can observe a user pointing at a specific image with the contextual interface rendered slightly above the controller and providing two options based on the current target of interest. This style of interaction was chosen based on evidence suggesting users prefer object-action sequences over action-object sequences as it requires less mental effort [16]. To navigate between these options the user uses their thumb on the controller's touchpad to highlight and make their selection. The context menu currently provides three possible functions depending on what image is being targeted. The first and most important function is 'Explore', which enables the user to explore all the images in an event. These explored images are presented in a line in front of the previously rendered results (see Figure 8) and can be navigated or scrolled through in an identical fashion by gripping the controller and performing a throwing gesture in the chosen direction.

The remaining two functions provided by the contextual interface are 'Zoom' and 'Search Tags' which are intended as secondary features provided to improve accessibility for some users. The 'Zoom' option significantly increases the scale of a target image to make it easier to examine. This is helpful in rare situations where the image is particularly detailed or contains a concept which may only occupy a small portion of the image. The 'Search Tags' option copies all of the concepts the target image is labelled with and reloads the main menu with those concepts prepared for submission. This is helpful when an image contains numerous relevant concepts and the user wants to quickly submit a related query.

4.5 Summary

In the preceding section we have discussed the methodology behind event visualisation within the latest iteration of the VRLE prototype. The generation and ranking of these events was based on the search engine developed by Duyen et al. [9]. With the introduction of event-based visualisations, we needed to introduce a process of keyframe selection to preview large events and also design an appropriate method by which the user could interact with these previews within the virtual environment.

5 CONCLUSION

In this paper we have presented our latest iteration of the VRLE prototype, an interactive retrieval system for lifelog data in virtual reality, distinguished from previous VRLE iterations by an emphasis on event-based data visualisation and a novel back-end search engine developed by Duyen et al. [9]. The participation of the VRLE prototype in the LSC 2020 is part of ongoing research to develop novel and effective techniques for interactive information retrieval in virtual reality within the context of personal data. In that respect our participation is intended to evaluate the effectiveness of our latest virtual data visualisation techniques and the appropriateness of the event ranking produced by our updated lifelog search engine.

REFERENCES

- [1] Matthew L. Cooper and Jonathan Foote. 2005. Discriminative techniques for keyframe selection. In *Proc. IEEE International Conference on Multimedia and Expo (ICME)*. IEEE Computer Society, Amsterdam, The Netherlands, 502–505.
- [2] Gabriel de Oliveira Barra, Alejandro Cartas Ayala, Marc Bolaños, Mariella Dimicoli, Xavier Giro i Nieto, and Petia Radeva. 2016. LEMore: A Lifelog Engine for Moments Retrieval at the NTCIR-Lifelog LSAT Task. In *Proc. NTCIR Conference on Evaluation of Information Access Technologies*. National Institute of Informatics (NII), Tokyo, Japan, 366–371.
- [3] Aiden R Doherty, Daragh Byrne, Alan F Smeaton, Gareth J.F. Jones, and Mark Hughes. 2008. Investigating keyframe selection methods in the novel domain of passively captured visual lifelogs. In *Proc. ACM International Conference on Image and Video Retrieval (CIVR)*. ACM, Niagara Falls, Canada, 259–268.
- [4] Aiden R. Doherty, Chris J. A. Moulin, and Alan F. Smeaton. 2011. Automatically assisting human memory: A SenseCam browser. *Memory* 19, 7 (2011), 785–795.
- [5] Aaron Duane. 2020. *Visual Access to Lifelog Data in a Virtual Environment*. Ph.D. thesis. Dublin City University.
- [6] Aaron Duane and Cathal Gurrin. 2017. Pilot Study to Investigate Feasibility of Visual Lifelog Exploration in Virtual Reality. In *Proc. Workshop on Lifelogging Tools and Applications (LTA@MM 2017)*. ACM, Mountain View, CA, USA, 29–32.
- [7] Aaron Duane and Cathal Gurrin. 2018. Lifelog Exploration Prototype in Virtual Reality. In *Proc. International Conference on Multimedia Modeling (MMM)*. Springer, Bangkok, Thailand, 377–380.
- [8] Aaron Duane, Cathal Gurrin, and Wolfgang Hürst. 2018. Virtual Reality Lifelog Explorer. In *Proc. ACM Workshop on Lifelog Search Challenge (LSC@ICMR 2018)*. ACM, Yokohama, Japan, 20–23.
- [9] Tran Ly Duyen, Nguyen Manh Duy, Nguyen Thanh Binh, Hyowon Lee, and Cathal Gurrin. 2020. Myscéal - An Experimental Interactive Lifelog Retrieval System for

- LSC'20. In *Proc. ACM Workshop on Lifelog Search Challenge (LSC@ICMR 2020)*. ACM, Dublin, Ireland.
- [10] Cathal Gurrin, Hideo Joho, Frank Hopfgartner, Liting Zhou, and Rami Albatat. 2016. Overview of NTCIR-12 Lifelog Task. In *Proc. NTCIR Conference on Evaluation of Information Access Technologies*. National Institute of Informatics (NII), Tokyo, Japan, 354–360.
 - [11] Cathal Gurrin, Hyowon Lee, Niamh Caprani, Zhenxing Zhang, Noel E. O'Connor, and Denise Carthy. 2010. Browsing Large Personal Multimedia Archives in a Lean-Back Environment. In *Proc. International Conference on MultiMedia Modeling (MMM)*. Springer, Chongqing, China, 98–109.
 - [12] Cathal Gurrin, Klaus Schoeffmann, Hideo Joho, Andreas Leibetseder, Liting Zhou, Aaron Duane, Duc Tien Dang Nguyen, Michael Riegler, Luca Piras, Minh-Triet Tran, Jakub Lokoč, and Wolfgang Hürst. 2019. Comparing Approaches to Interactive Lifelog Search at the Lifelog Search Challenge (LSC2018). *ITE Transactions on Media Technology and Applications* 7 (04 2019), 46–59.
 - [13] Alan Kwan. 2012. Bad Trip. <https://www.kwanalan.com/blank>
 - [14] Hyowon Lee, Alan F. Smeaton, Noel E. O'Connor, Gareth J. F. Jones, Michael Blighe, Daragh Byrne, Aiden R. Doherty, and Cathal Gurrin. 2008. Constructing a SenseCam visual diary as a media process. *Multimedia Systems* 14, 6 (2008), 341–349.
 - [15] Hsiang Lun Lin, Tzu-Chieh Chiang, Liang-Pu Chen, and Ping-Che Yang. 2016. Image Searching by Events with Deep Learning for NTCIR-12 Lifelog. In *Proc. NTCIR Conference on Evaluation of Information Access Technologies*. National Institute of Informatics (NII), Tokyo, Japan.
 - [16] Ryan P. McMahan and Doug A. Bowman. 2007. An Empirical Comparison of Task Sequences for Immersive Virtual Environments. In *IEEE Symposium on 3D User Interfaces (3DUI)*. IEEE Computer Society, Charlotte, NC, USA, 14.
 - [17] Marcus Jerome Pickering, Daniel Heesch, Stefan M. Rüger, Robert J. O'Callaghan, and David R. Bull. 2002. Video Retrieval Using Global Features in Keyframes. In *Proc. Text REtrieval Conference (TREC)*. National Institute of Standards and Technology (NIST), Gaithersburg, MD, USA.
 - [18] Zhengwei Qiu, Cathal Gurrin, and Alan F. Smeaton. 2016. Evaluating Access Mechanisms for Multimodal Representations of Lifelogs. In *Proc. International Conference on MultiMedia Modeling (MMM)*. Springer, Miami, FL, USA, 574–585.
 - [19] Bahjat Safadi, Philippe Mulhem, Georges Quénot, and Jean-Pierre Chevallet. 2016. LIG-MRIM at NTCIR-12 Lifelog Semantic Access Task. In *Proc. NTCIR Conference on Evaluation of Information Access Technologies*. National Institute of Informatics (NII), Tokyo, Japan.
 - [20] Harrisen Scells, Guido Zuccon, and Kirsty Kitto. 2016. QUT at the NTCIR Lifelog Semantic Access Task. In *Proc. NTCIR Conference on Evaluation of Information Access Technologies*. National Institute of Informatics (NII), Tokyo, Japan.
 - [21] Yang Yang, Hyowon Lee, and Cathal Gurrin. 2013. Visualizing lifelog data for different interaction platforms. In *Proc. ACM SIGCHI Conference on Human Factors in Computing Systems*. ACM, Paris, France, 1785–1790.
 - [22] Liting Zhou, Duc-Tien Dang-Nguyen, and Cathal Gurrin. 2017. A Baseline Search Engine for Personal Life Archives. In *Proc. Workshop on Lifelogging Tools and Applications (LTA@MM 2017)*. ACM, Mountain View, CA, USA, 21–24.