Myscéal: An Experimental Interactive Lifelog Retrieval System for LSC'20

Ly-Duyen Tran*
Manh-Duy Nguyen*
Dublin City University
Ireland

Hyowon Lee
Insight Centre for Data Analytics, Dublin City University
Ireland

Nguyen Thanh Binh
AISIA Research Lab
VNU HCM - University of Science
Vietnam

Cathal Gurrin
Dublin City University
Ireland

ABSTRACT

The Lifelog Search Challenge (LSC), is an annual comparative benchmarking activity for comparing approaches to interactive retrieval from multi-modal lifelogs. Being an interactive search challenge, issues such as retrieval accuracy, search speed and usability of interfaces are key challenges that must be addressed by every participant. In this paper, we introduce Myscéal, an interactive lifelog retrieval engine designed to support novice users to retrieve items of interest from a large multimodal lifelog. Additionally, we also introduce a new similarity measure called *"aTFIDF"*, to match a user's free-text information need with the multimodal lifelog index.

CCS CONCEPTS

• **Information systems** → **Information retrieval**; • **Human-centered computing** → **Human computer interaction (HCI)**; **User interface design**.

KEYWORDS

lifelog, interactive retrieval system, elasticsearch

1 INTRODUCTION

Although lifelogging has a long history [3], substantial progress has only been seen since the advent of the MyLifeBits project [7] and the related CARPE workshop [8] during the early years of this century. Such advances are a result of the ready availability of low-cost and power-efficient (mostly wearable) sensors which capture

---

*Both authors contributed equally to this research.

multi-modal digital evidences of daily activities, such as contextual images, audio, GPS, and personal biometric data like heart rate or weight [14]. These diverse data sources lead to the creation of huge personal data archives (called lifelogs) and these lifelogs pose a new challenge for the information retrieval community. Supporting personal lifelog search and retrieval is therefore an important research topic and has significant potential to act as an assistive technology [14]. Many competitions and workshops have been introduced to solve this challenge, such as the ImageCLEF workshops in recent years [25] and the NTCIR Lifelog tasks which ran between 2016 and 2019 and released the main datasets used by the community, such as [10]. The novel Lifelog Search Challenge (LSC) [13] builds upon these efforts by releasing data and coordinating the comparative evaluation of many interactive lifelog retrieval systems in an open, real-time and metrics-driven evaluation.

Interactive information retrieval systems are suitable for supporting access to large lifelogs [15]. Both LSC'18 and LSC'19 evaluated interactive systems of participants by asking research teams to find specific images related to queries within a time limitation. For each query, an additional hint can be provided every 30 seconds to imitate the real cases when people usually remember gradually over time. To perform well at the LSC challenge, a system must have an effective retrieval engine and must consider issues of usability, since a retrieval system will be tested separately by expert and novice users. In two years of organising the challenge, there are a variety of approaches for the searching system ranging from faceted windows in virtual reality [6], to interactive learning [16], and migrating proven approaches from video retrieval [18, 20, 28]. Initial findings suggest that an effective search mechanism coupled with an intuitive user interface are the key components of a competitive system [13].

In this paper, we introduce an interactive information searching system, named Myscéal (from the Gaelic word scéal meaning story), which has been developed specifically for novice (non-expert) users and utilises text querying as the main source of evidence for retrieval. A new ranking algorithm aTFIDF is introduced and we describe both visual and semantic features that were employed for metadata generation. We deploy the ElasticSearch back-end [9] for both storing and retrieving annotations generated from the LSC'20 dataset. Moreover, Myscéal system employs an automatic query expansion mechanism generated by a re-trained Word2Vec model [24], which serves to enhance the likelihood of query-to-index matching for novice users. In order to enhance novice usability, the

interface includes a (purposely) limited number of faceted searching options. Finally the system supports a novel query mechanism that facilitates sequence queries which allow the user to easily define a query that includes information about a sequence of user activities, which the authors felt would be critical for addressing the queries posed in previous LSC challenges.

2 RELATED WORK

The first interactive lifelog retrieval system, MyLifeBits [7], viewed lifelog retrieval as a database challenge. Bell & Gemmel proposed a simple interface supporting many view panels for the different media types. It was a system built for precise non-ranked inquiry. However, just like in the early days of web search, simple inquiry needed to be replaced with ranked lists as lifelogs became larger[14]. Recently, LEMoRe [5] was an early system designed specifically to support interactive lifelog retrieval, although there was no clear performance comparison at that time [11]. The concept behind LEMoRe was that images from personal events can be annotated with the semantic context to enrich the information and therefore make it easier to be later retrieved.

In 2018, the LSC contest was established to facilitate a fair comparison between different interactive lifelog retrieval systems. This is facilitated by comparing participating systems' performance on the same dataset and the same set of queries in a side-by-side real-time competition. The top ranked system in the first LSC was built on the virtual reality concept [6]. The interface of this system was straightforward for novice users with only visual concept and temporal filters. The result was visualized and ranked based on simple concept match frequencies. Although the interaction seems to be complicated as users need to handle two extra controllers and performed in the virtual environment, the intuitive and simple nature of the interface was key in its performance [13]. In each of the previous two LSC contests, many systems were based on existing interactive video search engines approaches and performed very well. LifeExplore [18], Viret [20], Vitrivr [28] all aggregated images within a day into a single video-type representation and created shots by grouping similar temporal images. Some also had options for querying by sketch. While LifeExplore focused on improving user interaction in the updated version for LSC'19 with filtering fields, Vitrivr introduced the Boolean query, which worked as a prior filter during the searching process. Viret could perform temporal queries that require retrieving two sequential moments. Other systems, such as LifeSeeker, [17], were built based on top of free-text search systems implementing textual ranking models such as BM25 [27]. The creators of Lifeseeker also also applied synonyms to expand text queries. Another notable system was Exquisitor [16] that could analyse the relevance feedback from users from which the system can retrieve relevant images without any input query by using interactive learning approach.

As with most of the prior systems, Myscéal is built on-top of a conventional text search engine. The primary contributions of Myscéal is the inclusion of an automatic query expansion mechanism, based on word-embedding, which assists novice users who do not know the indexed ontology for the collection. Additionally, a second contribution is a novel scoring function to contribute to the weights of the ranking called aTFIDF. Finally, we also developed

a simple but efficient user interface, which focuses on (natural) text queries, keeps the use of faceted search to a minimum [7] and supports the lifelog query triads of before, now, and after in a sequential query process designed specifically to support lifelog queries.

3 LSC'20 DATASET AND SCORING FUNCTION

Different from the previous LSC competitions, LSC'20 uses the personal data of a single lifelogger over four months between the years 2015 to 2018 [12]. Images in the dataset are redacted by blurring faces and sensitive textual information. Along with images, automatic visual concept annotations are also included. This consists of the results of scene detection and object detection performed on the original (non-redacted) visual data. Additionally, metadata, which includes information about times, location, activities and biometrics, is also provided.

The evaluation procedure is described in detail in [13], but we will mention briefly here because it's high importance given to correctness and speed of retrieval directly influence the decision decisions taken in our system development. Every searching system will be used by expert and novice users in 2 distinct sections. In each section, users will be given several tasks and each of successfully solved task will award participants the score of

$$S = max(0, A.\frac{T.0.9^\omega - 0.5.\tau}{T})$$

in which A is the maximum point for a task, T is the maximum allowed searching time and varies between expert and novice users, ω is the number of incorrect submission, and τ is the searching time of the user. This scoring function considers not only accurate retrieval, but also the searching time, which is influenced by the user interface of the systems.

4 MYSCÉAL SYSTEM

Myscéal follows the schema of a typical basic lifelog search engine [32] as depicted in Figure 1. Initially, we extract textual features from image annotations in the dataset of LSC'20, then create an inverted index in the underlying ElasticSearch engine. The GPS data from the locations in metadata is stored in a geospatial format, while the time information is indexed in a date format. Moreover, we also extend this information to enrich the indexed data, as described in the following section. Users can use a clear interface to perform interactive retrieval and results of any query are displayed in ranked order from which users can select and submit to the server. We will now describe each component of the system.

4.1 Feature extraction

The metadata provided can be split into four categories: location, time, activities, and visual concepts. Biometric data such as heart rate and calories were omitted because we believe they are not an essential factor in retrieval in most circumstances. For the visual concepts, besides the available ones (including place attributes, place categories, and visible objects), we also employ an additional object detector using the results of the semantic segmentation engine called DeepLabv3+ [4]. The neural networks were retrained on the ADE20k dataset [31]. However, we redefine the labels based on the similarities between some objects (for example, lamp and
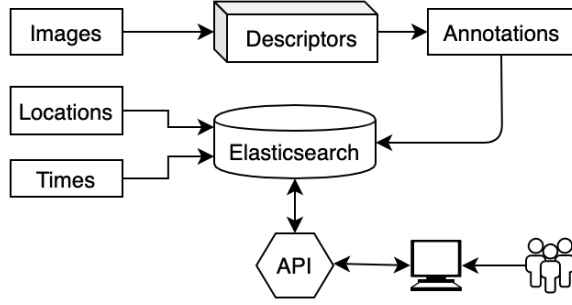
Figure 1: Myscéal system pipeline

light source). Additionally we include their bounding boxes in the segmentation maps of DeepLabv3. From the segmentation map, we calculate the area of objects based on the pixel counts and calculate a modified version of term frequency-inverse document frequency statistic (TF-IDF [29]), which reflects the relative importance of each annotated object and will be described in Section 4.4.

4.2 Data indexing and Enhancement

ElasticSearch is an open-source search and analytics engine built on Apache Lucene, which is a full-text search library [9]. This engine, however, can be used for retrieving other types of data such as images [1]. In this research, we index times data as dates format in ElasticSearch which supports any time-related queries. The GPS is stored as geospatial format which contains latitude and longitude information and can be retrieved by using the map area of our interface. On the other hand, the features created in section 4.1 will be indexed in original text format. We decided to enrich the database by enhancing times and GPS data. From time information, the day-of-the-week (Monday, Tuesday, ...) texts are created and can be searched as text format without the need for (unwieldy) faceted filters. The location text field is extended with the area names of the GPS points such as names of states and countries where the latitudes and longitudes are belong to. This will be useful when users do not know where the queried area is located in the map, hence they can search as a normal text retrieval. In this version, we have not used other data like music or heart rate, as this information did not contribute much to the two previous LSC events.

4.3 Query Expansion

Since our system does not support faceted search, users query using text. Hence, it is vital to classify words from the input query into corresponding fields in the ElasticSearch database. We create our rule-based part-of-speech tagging from the natural language toolkit library [2] to solve this issue. For instance, from the query "*having coffee at the helix at 9 am on 23 August*", the system will recognize that there is an object (coffee), location (the helix), time (9 am), and date (23 August) under our tagging and circulate them to the database for retrieving operation.

After we extract necessary fields of information in the part-of-speech tagging, the terms undergo another step in which they

would be mapped into a specific set of keywords that resulted from the feature extraction process. One can carry this step out by calculating the similarities between the input words and the keywords by a Word2Vec [24] model and WordNet [26] - an extensive lexical database of English. WordNet provides relationships among words such as synonyms (the user might enter *TV* and this would be mapped into the existing keyword *television*), hypernyms (a *child* is a *person*), hyponyms (a *fruit* might be an *apple*). The Word2Vec model is used to model the co-occurrence of words in one image. It was trained on the COCO detection and captioning dataset [19] as well as the LSC'20 dataset itself. The similarities calculated by the model are used to expand the entered word into similar concepts. For example, the word *office* suggests *computer* and *chair*.

4.4 Information Retrieval

There are two processes that influence the ranking of the images. The first simple case is based on filters. The filter will be automatically detected from the query input if there is an exact match. Following is the list of all possible filters and their examples:

- *location:* an exact name of the location (Dublin City University) or the region (Oslo, Norway).
- *activity:* driving, walking, airplane and transport.
- *time range:* before 6am, around 12:30pm, between 4pm and 6pm.
- *time of day:* this could be morning, lunch, late afternoon, etc.
- *date or day of week:* a specific date (May 15) could be used; Monday.
- *attributes and categories:* these visual labels are provided with the dataset.
- *GPS coordinates*: using the map filtering in Figure 3b

The expanded set of visual concepts resulted from section 4.3 is also used for filtering. We use the ElasticSearch "term sets" query in which more keywords matches result in a higher score. We call this score S_E.

However, the visual concepts detected in the feature extraction process are not suitable for using only filters because they, as well as the expanding process, are not sufficiently correct. In this case, an algorithm that is based on the ubiquitous TF-IDF term weighting approach is used to contribute to the ranking of the filtered set of images. We call this **aTFIDF** For a given image, if we denote the set of the images from the LSC'20 dataset by I, the collection of possible object keywords from section 4.1 by O, the area of an object detected in that image by $f_{o,i}$, $o \in O$, $i \in I$, we can calculate the **area-term frequency** as following:

$$aTF(o, i) = 1 + log(f_{o,i})$$

The **area-inverse document frequency** can be obtained by the following:

$$aIDF(o) = log\left(\frac{N}{|\{i \in I : f_{o,i} > c)\}|}\right)$$

where

- N: total number of images in the dataset
- c: a constant which is used as a threshold for the area for determining if an object is actually in the image or if it is visual noise.
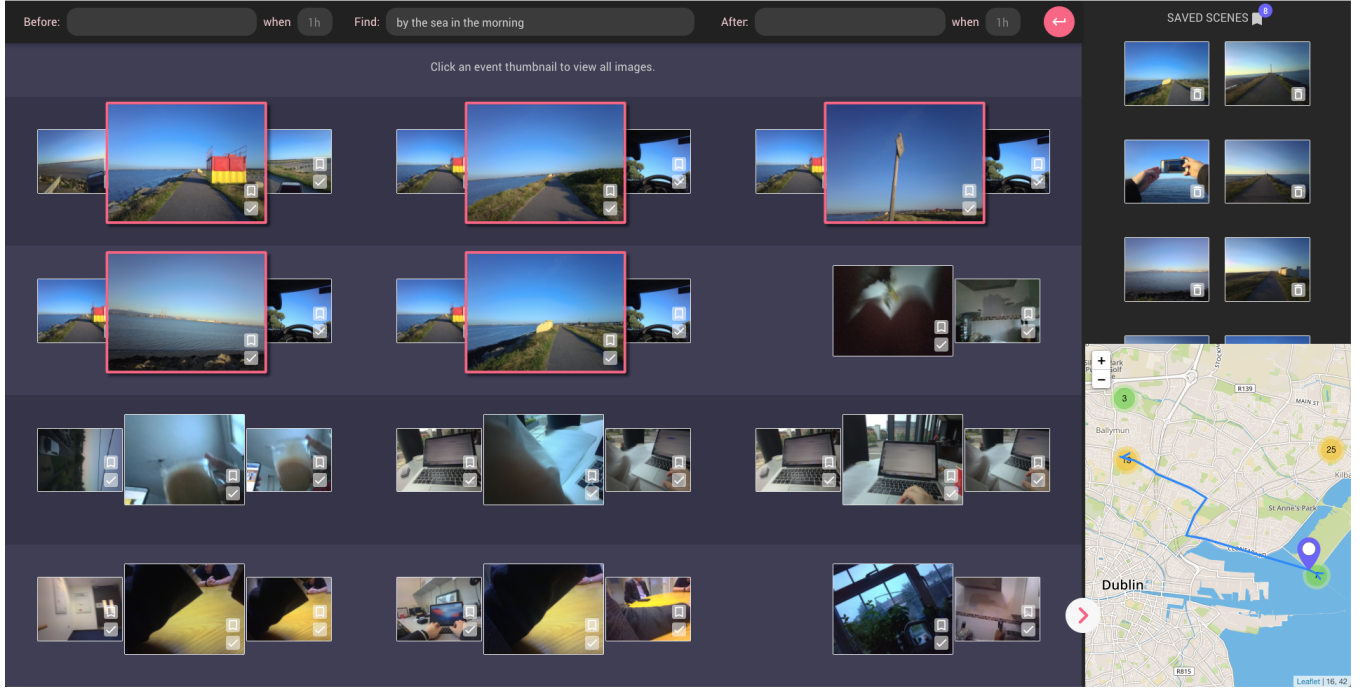
Figure 2: Myscéal User Interface

Finally, the **aTFIDF** can be calculated as follows:

$$aTFIDF(o, i) = aTF(o, i) * aIDF(o)$$

For a processed input query $q = [o_0, o_1, ..., o_m], o_j \in O$, the final score of each image can be formulated as:

$$S(q, i) = w_1 S_E + w_2 \sum_{j=0}^{m} aTFIDF(o, i)$$

where S_E is the score from the filtering process of ElasticSearch, w_1, w_2 is used for balancing the weights and required tuning.

4.5   Temporal Retrieval

Since lifelog data is naturally sequential and continuous in nature, the idea that a user would reference temporal relations in queries is obvious. In fact most of the previous LSC queries incorporated a temporal dimension, such as *I bought the bottle of wine after driving to the shop, before I went to the party*. For temporal retrieval to work, the data needs to be organised into logically related sequences of images and metadata, called events. Based on the redundancy of consecutive identical images in the dataset, such events could be identified. A change of event indicates an introduction of new information such as a change of action or a shift in point of view of the lifelogger. For instance, such changes could be when the lifelogger started to use the cellphone while watching television or looked up to see a colleague entering the office while using the laptop. In order to break the sequence of images into events, for each image, its visual features, which include SIFT [21–23] and a 4096-dimension vector from the pre-trained VGG16 model [30], were extracted along with the visual concept descriptions from section 4.4. All features are then compared with those of the immediately preceding image by using cosine distance to decide if they belong to the same event.

Myscéal can facilitate a temporal query containing up to three consecutive events. Considering the wine shopping query from the previous paragraph, firstly, the system will search for the main event (buying a bottle of wine). We extract the time information from the result and use it as additional input for the other two events (driving and attending a party). At the last step, the system can group all corresponding results and rank based on their total score. In order to make the result becomes more informative, the temporally related two events are extended on screen to include several extra events within a given period of time. They will be shown when the user clicks on the thumbnail as seen in the small row below in Figure 3a.

4.6   User Interaction

The main challenge in the design of the user-interface was to integrate the back-end retrieval and query expansion functions in such a way as to optimally assist the user in the particular task scenario defined in the LSC. Efficiency in the task completion, i.e. searching for the photo as quickly as possible, was set as the usability priority, and all UI features were designed accordingly. However, considerations on the novice users had to be also taken into account as the competition includes the first-time users to be recruited and use the system with only a brief introduction/training at the start. Considering these, the overall strategy developed and incorporated in the UI are to:
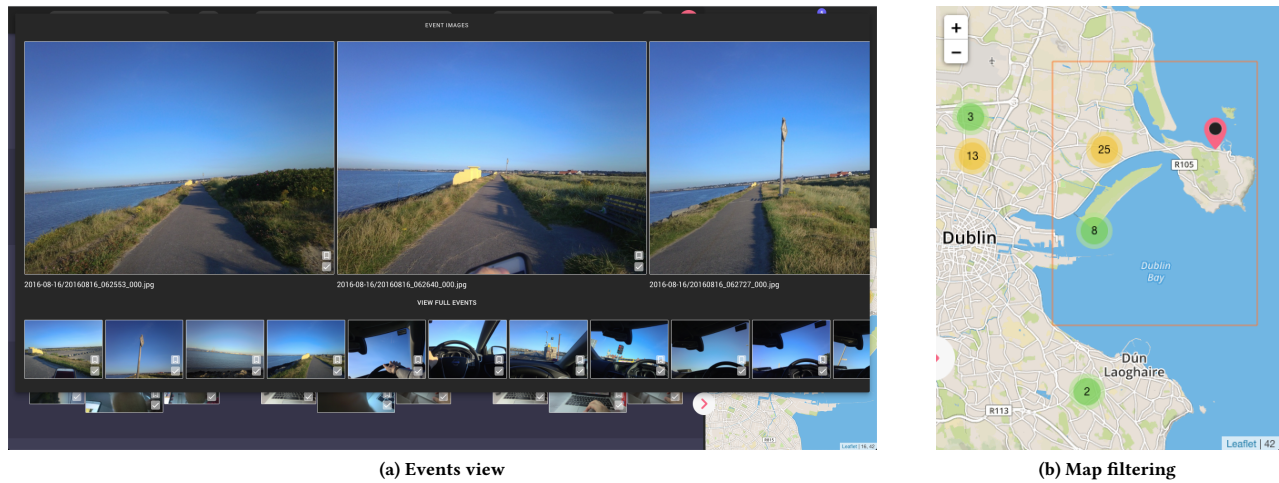
(a) Events view

(b) Map filtering

Figure 3: Faceted windows in Myscéal

- minimise different steps/stages in user's search process: although the generic information search process has been understood well and often the search UIs follow these processes, the most critical step in the task scenario of this challenge needs to be identified and avoid temporal jumping from one screen to another in order to minimise the search time;
- find ways for the back-end functionalities to be the guiding feature for interactivity: for the indexing and processing to best support the end-user search, its strengths must be made full use of, in the way the user conducts the task. For example, the temporally-designated retrieval (Section 4.4) means the user can indicate the query elements to be either before or after the searched event in concern: the UI should represent this clearly so that the user can make use of it readily.

Figure 2 shows a screenshot of the designed UI. As the first part of the query is displayed, the user starts by typing in at the search bars (top of the screen) with three separate query boxes where temporal relationships amongst the query parts (before, after, and the one to be be found) can be expressed. The reason behind these additional boxes is that while our part-of-speech tagging could capture the temporal feature of a single event, but not between different events. Determining which of the boxes the query should be entered will be straightforward and can be quickly done, as the query portion appearing will indicate whether it is the event, or something that happened before or after the event to be searched. The user can specify the timing of "before" and "after" queries by entering the number of hours before/after the main event being sought (small boxes between the three main query boxes).

As soon as entering query terms in either of the 3 boxes, the retrieval result is displayed in the main part of the screen as a vertically-scrollable panel. Each entry in the retrieval result is again guided corresponding to the way query terms were entered: the event to be searched is represented by the photo in the middle, and the events before and after as the retrieval matched are represented by two smaller photos on the left and right side. As each photo represents an event, there could be multiple photos included in it. By clicking any of the photos in this result, all photos belonging to the event appears on top of the main screen (as illustrated in Figure 3a) until other parts of the screen is clicked. The temporary nature of the appearance of these within-event photos means there is no major refresh or transitions of the screen estate as far as the user is concerned, reducing any small amounts of time for mental re-orientation typically required after such a transition.

As the query is progressively revealed to the user, the UI offers a "Saved scenes" panel (on top-right of the screen depicted in Figure 2) where the user can temporarily save any photos and events found in any stages of the revealed query. Any candidate photo(s) the user thinks could be correct can be saved here without submitting, while continuing search and wait until the full query details have not been revealed. Below the panel, a geographic map is shown as part of retrieval result by default and the locations as matched by the system indicated and centred. Figure 3b shows another feature of the map which also help faster search in certain queries where the visualisation of location is a useful information. Users can draw a rectangle on the map to perform searching within that area (e.g. beach scene or a park). The indicated locations on the map will be highlighted in red as the user moves the mouse cursor over the entries in the search result, and vice versa. If the user decides the map is not a useful feature in the query, s/he can click on the arrow button at a lower part of the map to slide out the map and claim more space for browsing the retrieval result. At any time of the interaction, the user can move the mouse cursor over a photo to see a 'submit' button and clicking on it will submit the photo as the answer to the query.

In this way, the UI is "anchored" on the main screen as shown in the figure, while temporarily linked event photos pops up, map slides in/out, and the 'submit' button appears/disappears as the user moves the cursor around the screen. We expect this will effectively cater for the time-pressed, efficient search situation as well as supporting the learnability for the first-time users without sacrificing the efficiency for the expert users.

5 CONCLUSION

In this paper, we have introduced our first interactive lifelog retrieving system to compete in the LSC'20 event and the new "aT-FIDF" algorithm for ranking documents based on its visual features. Myscéal is built to address three crucial challenges, which are accurate searching, fast processing, and straightforward but high efficient design by deploying the ElasticSearch engine combined with an automatic query suggestion procedure. The system supports querying sequential moments and visualising the movements between them on the map. This map can work as a filtering option also.

Although the most valuable information used in the retrieving process is a visual concept, our system has not included other data such as heart rate or music. In the future, we might take them into account to utilize all given elements in the dataset. Besides, visual similarity retrieving is also intriguing to be investigated in the next Myscéal version.

ACKNOWLEDGMENTS

REFERENCES

[1] Giuseppe Amato, Paolo Bolettieri, Fabio Carrara, Fabrizio Falchi, and Claudio Gennaro. 2018. Large-scale image retrieval with elasticsearch. In *The 41st International ACM SIGIR Conference on Research & Development in Information Retrieval.* 925–928.

[2] Steven Bird, Ewan Klein, and Edward Loper. 2009. *Natural language processing with Python: analyzing text with the natural language toolkit.* " O'Reilly Media, Inc.".

[3] Vannevar Bush. 1996. As We May Think. *Interactions* 3, 2 (March 1996), 35–46. https://doi.org/10.1145/227181.227186

[4] Liang-Chieh Chen, Yukun Zhu, George Papandreou, Florian Schroff, and Hartwig Adam. 2018. Encoder-Decoder with Atrous Separable Convolution for Semantic Image Segmentation. In *ECCV.*

[5] Gabriel de Oliveira Barra, Alejandro Cartas Ayala, Marc Bolaños, Mariella Dimiccoli, Xavier Giró Nieto, and Petia Radeva. 2016. Lemore: A lifelog engine for moments retrieval at the ntcir-lifelog lsat task. In *Proceedings of the 12th NTCIR Conference on Evaluation of Information Access Technologies.*

[6] Aaron Duane, Cathal Gurrin, and Wolfgang Huerst. 2018. Virtual reality lifelog explorer: lifelog search challenge at ACM ICMR 2018. In *Proceedings of the 2018 ACM Workshop on The Lifelog Search Challenge.* 20–23.

[7] Jim Gemmell, Gordon Bell, and Roger Lueder. 2002. MyLifeBits: fulfilling the Memex vision. *Proceedings of the tenth ldots* (2002), 235–238.

[8] Jim Gemmell, Lyndsay Williams, Ken Wood, Roger Lueder, and Gordon Bell. 2004. Passive capture and ensuing issues for a personal lifetime store. In *Proceedings of the the 1st ACM workshop on Continuous archival and retrieval of personal experiences.* 48–55.

[9] Clinton Gormley and Zachary Tong. 2015. *Elasticsearch: the definitive guide: a distributed real-time search and analytics engine.* " O'Reilly Media, Inc.".

[10] Cathal Gurrin, Hideo Joho, Frank Hopfgartner, Liting Zhou, and Rami Albatal. 2016. NTCIR Lifelog: The First Test Collection for Lifelog Research. *Proceedings of the 39th International ACM SIGIR conference on Research and Development in Information Retrieval - SIGIR '16* (2016), 705–708. https://doi.org/10.1145/2911451.2914680

[11] Cathal Gurrin, Hideo Joho, Frank Hopfgartner, Liting Zhou, Rashmi Gupta, Rami Albatal, Dang Nguyen, and Duc Tien. 2017. Overview of ntcir-13 lifelog-2 task. NTCIR.

[12] Cathal Gurrin, Tu-Khiem Le, Van-Tu Ninh, Duc-Tien Dang-Nguyen, Björn Þór Jónsson, Jakub Lokoč, Wolfgang Hurst, Minh-Triet Tran, and Klaus Schoeffmann. 2020. An Introduction to the Third Annual Lifelog Search Challenge, LSC'20. In *ICMR '20, The 2020 International Conference on Multimedia Retrieval.* ACM, Dublin, Ireland.

[13] Cathal Gurrin, Klaus Schoeffmann, Hideo Joho, Andreas Leibetseder, Liting Zhou, Aaron Duane, Dang Nguyen, Duc Tien, Michael Riegler, Luca Piras, et al. 2019. Comparing approaches to interactive lifelog search at the lifelog search challenge (LSC2018). *ITE Transactions on Media Technology and Applications* 7, 2 (2019), 46–59.

[14] Cathal Gurrin, Alan F Smeaton, Aiden R Doherty, et al. 2014. Lifelogging: Personal big data. *Foundations and Trends® in information retrieval* 8, 1 (2014), 1–125.

[15] Diane Kelly et al. 2009. Methods for evaluating interactive information retrieval systems with users. *Foundations and Trends® in Information Retrieval* 3, 1–2 (2009), 1–224.

[16] Omar Shahbaz Khan, Björn Þór Jónsson, Jan Zahálka, Stevan Rudinac, and Marcel Worring. 2019. Exquisitor at the lifelog search challenge 2019. In *Proceedings of the ACM Workshop on Lifelog Search Challenge.* 7–11.

[17] Tu-Khiem Le, Van-Tu Ninh, Duc-Tien Dang-Nguyen, Minh-Triet Tran, Liting Zhou, Pablo Redondo, Sinead Smyth, and Cathal Gurrin. 2019. LifeSeeker: Interactive Lifelog Search Engine at LSC 2019. In *Proceedings of the ACM Workshop on Lifelog Search Challenge.* 37–40.

[18] Andreas Leibetseder, Bernd Münzer, Manfred Jürgen Primus, Sabrina Kletz, Klaus Schoeffmann, Fabian Berns, and Christian Beecks. 2019. lifeXplore at the lifelog search challenge 2019. In *Proceedings of the ACM Workshop on Lifelog Search Challenge.* 13–17.

[19] Tsung-Yi Lin, Michael Maire, Serge J. Belongie, Lubomir D. Bourdev, Ross B. Girshick, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C. Lawrence Zitnick. 2014. Microsoft COCO: Common Objects in Context. *CoRR* abs/1405.0312 (2014). arXiv:1405.0312 http://arxiv.org/abs/1405.0312

[20] Jakub Lokoč, Tomáš Souček, Premysl Čech, and Gregor Kovalčík. 2019. Enhanced VIRET tool for lifelog data. In *Proceedings of the ACM Workshop on Lifelog Search Challenge.* 25–26.

[21] D. G. Lowe. 1999. Object recognition from local scale-invariant features. In *Proceedings of the Seventh IEEE International Conference on Computer Vision*, Vol. 2. 1150–1157 vol.2.

[22] David G. Lowe. 2004. Distinctive Image Features from Scale-Invariant Keypoints. *International Journal of Computer Vision* 60, 2 (Nov 2004), 91–110. https://doi.org/10.1023/B:VISI.0000029664.99615.94

[23] H. Luo, H. Wei, and L. L. Lai. 2011. Creating Efficient Visual Codebook Ensembles for Object Categorization. *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans* 41, 2 (2011), 238–253.

[24] Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. 2013. Distributed representations of words and phrases and their compositionality. In *Advances in neural information processing systems.* 3111–3119.

[25] Dang Nguyen, Duc Tien, Michael Riegler, Liting Zhou, and Cathal Gurrin. 2018. Challenges and opportunities within personal life archives. (2018).

[26] Peter Oram. 2001. WordNet: An electronic lexical database. Christiane Fellbaum (Ed.). Cambridge, MA: MIT Press, 1998. Pp. 423. *Applied Psycholinguistics* 22, 1 (2001), 131–134. https://doi.org/10.1017/S0142716401221079

[27] Stephen E Robertson and Karen Spärck Jones. 1994. *Simple, proven approaches to text retrieval.* Technical Report. University of Cambridge, Computer Laboratory.

[28] Luca Rossetto, Ralph Gasser, Silvan Heller, Mahnaz Amiri Parian, and Heiko Schuldt. 2019. Retrieval of structured and unstructured data with vitrivr. In *Proceedings of the ACM Workshop on Lifelog Search Challenge.* 27–31.

[29] Gerard Salton and Michael J McGill. 1986. Introduction to modern information retrieval. (1986).

[30] Karen Simonyan and Andrew Zisserman. 2014. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556* (2014).

[31] Bolei Zhou, Hang Zhao, Xavier Puig, Sanja Fidler, Adela Barriuso, and Antonio Torralba. 2017. Scene Parsing through ADE20K Dataset. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.*

[32] Liting Zhou, Duc-Tien Dang-Nguyen, and Cathal Gurrin. 2017. A baseline search engine for personal life archives. In *Proceedings of the 2nd Workshop on Lifelogging Tools and Applications.* 21–24.