E-Myscéal: Embedding-based Interactive Lifelog Retrieval System for LSC'22

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

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

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

Cathal Gurrin
Dublin City University
Ireland

ABSTRACT

Developing interactive lifelog retrieval systems is a growing research area. There are many international competitions for lifelog retrieval that encourage researchers to build effective systems that can address the multimodal retrieval challenge of lifelogs. The Lifelog Search Challenge (LSC) was first organised in 2018 and is currently the only interactive benchmarking evaluation for lifelog retrieval systems. Participating systems should have an accurate search engine and a user-friendly interface that can help users to retrieve relevant content. In this paper, we upgrade our previous Myscéal, which was the top performing system in LSC'20 and LSC'21, and present E-Myscéal for LSC'22, which includes a completely different search engine. Instead of using visual concepts for retrieval such as Myscéal, the new E-Myscéal employs an embedding technique that facilitates novice users who are not familiar with the concepts. Our experiments show that the new search engine can find relevant images in the first place in the ranked list, four a quarter of the LSC'21 queries (26%) by using just the first hint from the textual information need. Regarding the user interface, we still keep the simple non-faceted design as in the previous version but improve the event view browsing in order to better support novice users.

CCS CONCEPTS

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

KEYWORDS

lifelog, interactive retrieval system, human factors

ACM Reference Format:
Ly-Duyen Tran, Manh-Duy Nguyen, Binh Nguyen, Hyowon Lee, Liting Zhou, and Cathal Gurrin. 2022. E-Myscéal: Embedding-based Interactive Lifelog Retrieval System for LSC'22. In *Proceedings of the 5th Annual Lifelog Search Challenge (LSC '22), June 27–30, 2022, Newark, NJ, USA.* ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3512729.3533012

*Both authors contributed equally to this research.

LSC '22, June 27–30, 2022, Newark, NJ, USA

ACM ISBN 978-1-4503-9239-6/22/06.
https://doi.org/10.1145/3512729.3533012

1 INTRODUCTION

Lifelogs are a form of personal data archive that captures a rich digital trace of the daily activities of an individual [13]. A lifelog dataset can contain many types of data from various sources, such as PoV visual images, semantic locations, or biometric signals [1, 8, 32]. Passively captured lifelog archives grow large very fast and require effective organisation and retrieval tools to be useful for the individual. Consequently, due to the ready availability of lifelog data sources, the research community has noticed and taken on the challenge. Many international workshops and activities have been organised and a first generation of lifelog dataset is released to facilitate comparative evaluation [10–12, 22]. Among them, the Lifelog Search Challenge (LSC) [12] was the pioneer in creating an interactive benchmark evaluation to assess the performance of the participating systems in real time. Participating systems compete in solving tasks where each of the tasks requires finding a lifelog image within the dataset that is relevant to a given textual query. This query is gradually revealed with additional information every 30 seconds. After 5 iterations, participants will have a final 150 seconds to find the answer, meaning that they will have to find the correct images in 300 seconds in total. This revealing mechanism simulates human recall of an event from the past, where memory becomes clearer as it is considered in more detail. After successfully solving the task, the system will be awarded a score which is based on the time it took to find the relevant item, if found at all. Therefore, a lifelog retrieval system attending LSC should have an accurate and fast search engine to be able to find the correct answer in a short amount of time.

Myscéal was the LSC winning system in the two recent LSC'20 and LSC'21 [35, 36]. Although Myscéal was one of the teams that solved the most queries, the main factor that helped Myscéal obtain the highest score compared to other systems was the prompt retrieval time. Fast retrieval was facilitated by the concept-based search engine and the straightforward user interface. Due to the pandemic, both LSC'20 and LSC'21 only facilitated expert users, so there was no opportunity to hold a novice session where each system will be used by a novice user who is not familiar with the system. In anticipation of an in-person event in 2022, the new version of Myscéal developed for LSC'22 is called E-Myscéal which

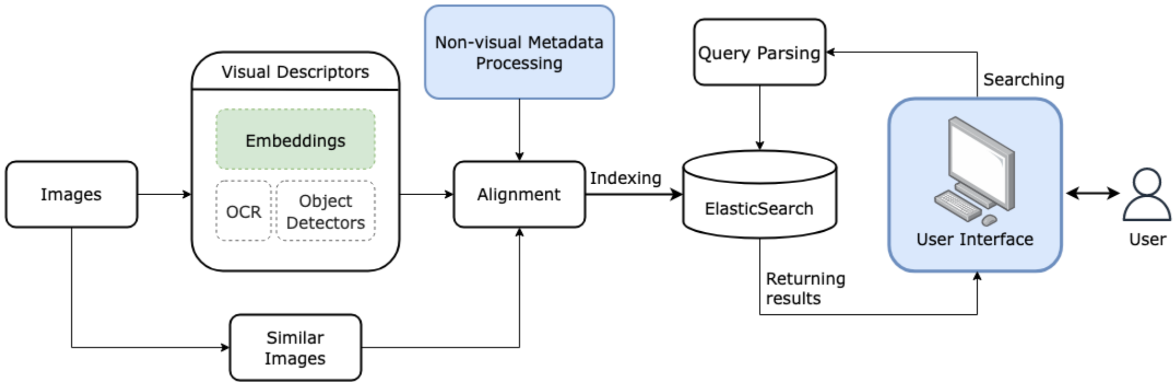

Figure 1: The E-Myscéal system pipeline keeps the general design from previous versions with adjustments. The components highlighted in blue are modified and improved. The embeddings (in green) are newly added, giving the new system its name.

introduces a new internal search engine and an easy-to-use user interface, which is better suited to novice users.

2 RELATED WORK

Because of the large size of lifelog datasets, it is important to have a retrieval system that can produce a ranked list rather than a conventional database set-based mechanism [13]. One of the pioneering systems that adopted this idea was LeMore [5], an interactive lifelog retrieval system from 2015. The pipeline of this system was to annotate each lifelog image with the textual information that is extracted from each image. This semantic detail can be the objects appearing in the image, the location indicating where the image was taken, or the activity context. In doing this, the visual images were converted into a textual modality, where they could be compared with a textual query to generate a ranked list. This effective method has been applied to winning systems of the LSC challenge since the inception of the event [6, 29, 35, 36].

There have been many lifelog retrieval systems that participated in the LSC competition with a variety of search methods. The most common approach is to compare the keywords in the queries with the annotated concepts of lifelog images to find the answer. Although many teams utilised this idea [4, 15, 25, 34, 36], each implements their own unique features. Lifeseeker [25] indexed the annotated information of the images with Elasticsearch [9] that supports text retrieval using a conventional TF-IDF [30] retrieval approach. Despite using the same database, Myscéal [36] applied a unique aTF-IDF scoring function in addition to the conventional one. Similarly to TF-IDF, the aTF-IDF function used in Myscéal was based on the frequency of the semantic annotated concepts in the database. However, this function also took into consideration the area of concepts, which were objects appearing in the images. Both Myscéal and LifeConcept [4] were integrated with a concept recommendation technique to address the gap between textual queries and visual annotated concepts. Because the lifelog dataset contains a large number of blurry images due to the physical activities of the lifelogger, vitrivr [15] employed a stabilising model. Meanwhile, vitrivr-VR [34] shared the same engine as vitrivr, but was developed for a virtual environment, which was an approach of the winning system in the first LSC challenge [6]. Another VR-based system, called VIRMA [7], created a hierarchy structure of visual concepts that supports users to explore the lifelog data rather than retrieval. Some systems applied the graph structure to lifelog dataset [24, 27]. Lifegraph [28] used the knowledge graph to enhance the annotated information. XQC[18] was one of the first lifelog retrieval systems designed to be used on a mobile device. The system relies on the user's relevance feedback through a number of iterations to gradually find the result, with the addition of a faceted filtering mechanism. This was also the approach of Exquisitor [16] in LSC'21. Due to an increase in the performance of embedding models in the image retrieval field [19, 23, 26], many lifelog retrieval systems are now applying this approach [2, 3, 20, 37]. Memento [2] and Voxento [3] were two of the teams that achieved high performance in LSC'21. Both systems used the same search engine which was built based on the CLIP model [26] from OpenAI. However, the latter system supports voice control, helping searchers interact with the system by using voice commands. SOMHunter+ [20] initially used W2VV++ model [19] to find the images that match the query, then revised the results based on the relevant feedback from the users.

In this paper, we propose to change the concept-based search engine in Myscéal [35, 36] to the embedding-based approach and present E-Myscéal (**E** for **E**mbedding). This modification is expected to facilitate novice users who do not have knowledge of the annotated concepts. We conducted an experiment to evaluate the effectiveness of the new search algorithm. In addition, a new interface is also proposed. Some confusing areas in the previous design have been removed to make E-Myscéal cleaner with fewer components to show to users. Therefore, our main contributions are threefold. First, we replace the search engine of Myscéal with a new embedding-based approach to introduce the entire E-Myscéal which is completely different from its original version. Second, we evaluate the performance of the new search algorithm with the experiment using the LSC'21 queries. Finally, we update the user interface with fewer components but additional guidelines to help users operate E-Myscéalwith ease.

3 THE LSC'22 DATA

Unlike LSC'21, LSC'22 will use a new lifelog dataset [14]. It is still a personal dataset gathered by the same lifelogger as in previous years. However, the multimodal dataset used in this year will be 18 months in length compared to 4 months as in LSC'20 and LSC'21. This results in more than 725K images in the archive. The bigger dataset requires retrieval systems to have a more powerful search engine to ensure a competitive search time. All sensitive information and faces in all images are anonymised to ensure privacy requirements. Each image is provided along with other lifelog information such as biometrics, sleeping state, date, and GPS signals indicating where and when the image was taken. In addition, some visual concepts and captions that describe the context of images are provided by Microsoft Computer Vision API[1]. Because OCR (information on text visible in an image) played a critical role in LSC'21, the dataset in LSC'22 comes with the extracted OCR details from Google Cloud Vision API[2].

4 OVERVIEW OF THE MYSCÉAL SYSTEM

In this section, we will briefly describe Myscéal in previous years. The pipeline of Myscéal has remained mostly unchanged since the first version, illustrated in Figure 1.

Data processing components provide each image with its associated visual descriptors and non-visual metadata, namely, GPS coordinates, semantic location names, time, and date. In addition, similar images are also computed at this stage using a combination of VGG16 [33] and SIFT [21] features. After that, all images are indexed in ElasticSearch, an open-source full-text search library, to enable high-speed search operations.

As Myscéal's original search approach was concept-based, the visual concepts obtained from the query often do not match with the indexed "keyword" concepts from object detectors. To overcome this, Myscéal employs a query expansion mechanism to modify each visual concept into a set of available keywords. This can be seen in the Query Parsing in Figure 1.

User interactions with the system are performed on a user interface as seen in Figure 2. In principle, Myscéal aims to support a novice user by minimising interaction steps. This motivates the decision to use a full-text search instead of relying on a faceted filter panel. The necessary information, both non-visual metadata and visual descriptors, can be directly parsed from the textual query by a customised query interpreter. Furthermore, Myscéal also supports searching for multiple queries based on their temporal relationships, which is a distinguishing characteristic of lifelogs. This is shown by the design of three separate query boxes (before, during, and after) at the top of the interface, and furthermore reinforced by showing the search results in triplets, putting the images in their temporal context.

Occasionally, location information cannot be well represented in text. Some examples could be "I was walking *near the coast*" or "I am running in a park *near my house*". If such instances occur, the user can locate the target location and perform filtering using the map panel on the top right. Moreover, the saved section allows the user to put aside images that they are not yet certain for later consideration.

[1] https://azure.microsoft.com/en-us/services/cognitive-services/computer-vision/
[2] https://cloud.google.com/vision

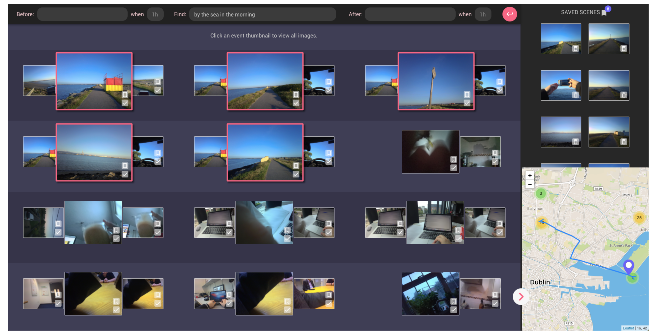

Figure 2: Myscéal user interface [35]

5 E-MYSCÉAL

5.1 Data processing

With the release of a new lifelog dataset for this year's challenge, some raw data need processing was performed to maintain some crucial search operations. In particular, the dataset lacks some rich semantic location annotations. This is essential, as in the past years, most queries contain location information, some of which are not available solely from the address provided by a geocoding reverse lookup. To handle this, we run a location clustering process inspired by Kikhia et al.[17]. The common approach to discovering location data patterns is to use density-based clustering algorithms [31], considering GPS coordinates and accompanying timestamps as input. From the resulting clusters, we can identify possibly important locations. Moreover, instead of using simple circles around the cluster center as we did last year, we construct a convex hull over each cluster to give an estimated boundary of the place. The name of the cluster is derived from the most common semantic name of its data points. From the boundaries, we can fill in missing or noisy semantic names for the rest of the dataset.

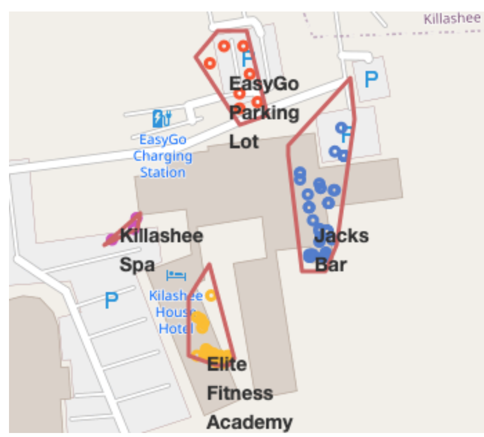

Figure 3: Examples of location clusters.

Furthermore, following Alam et al.[2], we remove unhelpful images from the dataset to reduce the size of the dataset and improve the performance of our event segmentation process. Unhelpful images are blurry, obstructed (by the lifelogger's arms, hands, or other objects), or dark images. By extracting SIFT features from all images, those without detectable SIFT features are considered unhelpful, as shown in Figure 4. As a reminder, Myscéal segments lifelog data into *scenes* as lifelog units, based on the combination of VGG16 and SIFT features. An example of images that are grouped into a *scene* can be seen in the right pane of the pop-up in Figure 5. In the presence of unhelpful images, a *scene* can be overly segmented, defeating the initial purpose of segmentation.

5.2 Embedding-based search approach

Our own experience with the system in previous challenges leads us to believe that the concept-based search, where we associate each image with tags for the detected objects (including their areas calculated as their total of pixels in the images), while sufficient on many occasions, fails to connect the semantic meaning of the query to images.

Due to the robustness of state-of-the-art embedding models, we decide to change our search approach from concept-based to embedding-based, utilising CLIP models from OpenAI [26]. CLIP models, which are joint models of vision and natural language, learn to perform a wide range of tasks and allow zero-shot transfer to many existing datasets. This model proved to be directly applicable to lifelog data as used in some participant systems in LSC'21, namely, Memento [2] and Voxento [3]. However, zero-shot CLIP fails to generalise to truly out-of-distribution data [26]. For this reason, we continue to use OCR results from Google Vision API in the same manner as in the old version.

Figure 4: Examples of unhelpful images.

Facilitating CLIP models renders the use of object detectors in previous versions of Myscéal obsolete. Therefore, the aTFIDF properties are not computed except for the OCR letters (as explained in the preceding paragraph). Additionally, as we could not deny the usefulness of the visual concepts provided with the dataset, we continue to use them as filters to reduce the computational space.

5.3 User Interface

The user interaction of the system has been revised, refined, and improved to better support novice users, as summarised now.

The triplet display layout in previous versions showed the retrieval results in *scenes* with their temporal context, taking into account the special characteristics of the lifelog data the system presents, i.e., the chronology of lifelog photos in terms of how the user could formulate the query and how the retrieval mechanism exploits it. However, when the query does not pertain to the temporal aspects, the triplet layout is not meaningful and only reduces the screen estate where more matched scenes could be accommodated. The refined retrieval result display flexibly changes the layout depending on the query: It was decided that the benefit of seeing more match result scenes without the "before-now-after" type of query outweighs the drawback of potential confusion in seeing two different layouts in the retrieval result display, especially in the time-constrained search context in LSC. Furthermore, as the two different retrieval result layouts are a direct consequence of either entering the query text in one or more query text boxes, a few times of querying is expected to naturally inform the user where the difference comes from.

When the user clicks on one of the retrieved scenes in either layout, a pop-up window appears (see Figure 5) that helps the user take further steps to either:

- *browse temporally-nearby scenes* in the selected day, with the selected scene highlighted in red: Being able to navigate through scenes around the scene selected in the retrieval result is a useful feature in the lifelog retrieval context, as more similar scenes might appear around the scene. The nearby scenes are vertically organised in groups and labelled as the result of the embedding-based modelling described in the previous section, shown in Figure 5 in a blue textual guideline on the vertical timeline.
- *browse other visually-similar scenes* to the selected scene, extracted from the whole archive: this feature is a secondary retrieval mechanism as the user browses through the chronologically arranged scenes above, increasing the chance of the user identifying the relevant scenes while browsing the within-day scenes.

On the right side of the pop-up window (the left of the geographical map) is a vertical panel with enlarged photos belonging to the selected scene, to support fast browsing within a scene that contains multiple images in it. Moving a mouse cursor over the small icon at the bottom right corner of each photo will magnify the size of the photo further to help with a more detailed inspection, consistent with the behaviour of the photos in all other panels (within-day scene list panel, "similar scenes" panel, and the initial search result screen). The user can close the pop-up window to return to the initial search result screen by pressing the ESC key on the keyboard.

Limiting the stages of search interaction to the retrieval result display and the pop-up window, along with the added labels and short instructions for each panel, is expected to help novice users. Balancing this with supporting efficient locating of potentially relevant images through within-day navigation along with similar scenes and enlarged within-scene images was the key consideration in this round of revision of the interface.

6 EVALUATION USING LSC'21 QUERIES

Despite our efforts to automatically parse the queries to reduce the discrepancy between an expert user and novice users, our system still requires human interpretation to split the information into "before", "main", and "after" queries. In this section, we will analyse the performance of E-Myscéal on the queries used in LSC'21. From this, we measure the Recall at K ($R@K$) from the results, ignoring further actions such as map filtering, temporal browsing, or visual similarity search. $R@K$ indicates whether one of the answer images appears in the first K images returned by E-Myscéal.

It has been mentioned that each query in LSC will be gradually revealed to the participants every 30 seconds. These additional hints

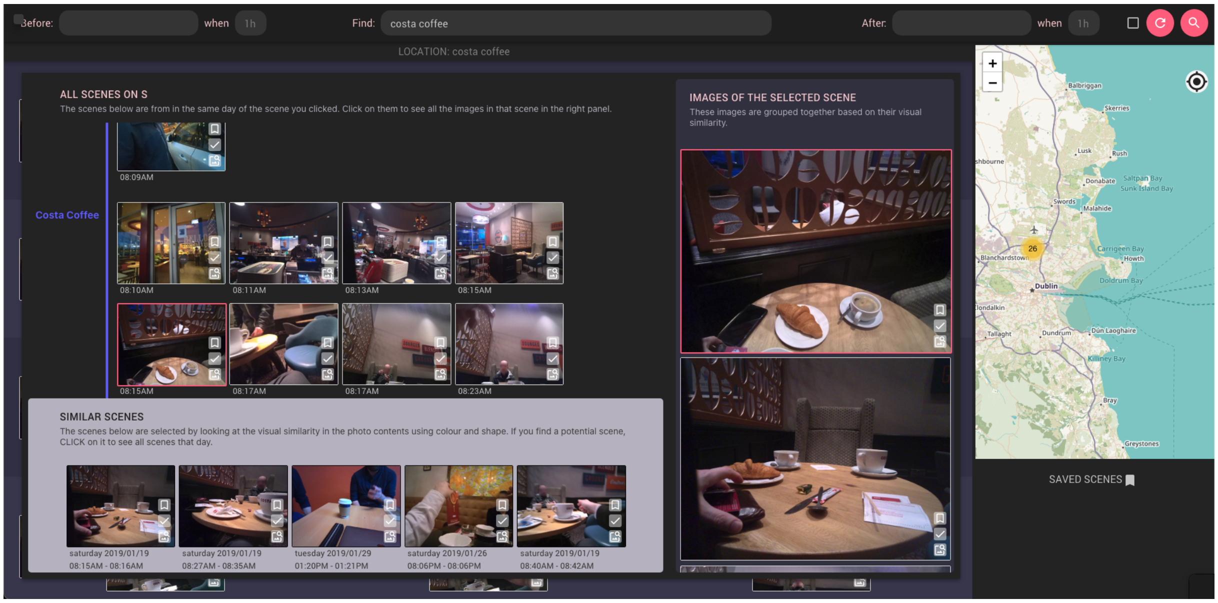

Figure 5: Event view window.

Table 1: Mean $R@K$ for LSC'21 queries.

Hint	R@1	R@3	R@5	R@10	R@20	R@50	R@100
1	0.26	0.35	0.39	0.52	0.52	0.52	0.57
2	0.30	0.35	0.43	0.48	0.57	0.57	0.57
3	0.35	0.43	0.52	0.57	0.57	0.57	0.57
4	0.26	0.43	0.52	0.57	0.57	0.57	0.57
5	0.35	0.48	0.52	0.61	0.61	0.65	0.65
6	0.48	0.52	0.61	0.61	0.61	0.61	0.61

are expected to make the query easier by offering more information about the content of the correct images. This results in a query consisting of 6 hints in total. Therefore, we evaluate E-Myscéal with the number of hints that are used to solve a query. Table 1 shows the average $R@K$ of E-Myscéal. We can see that our new system can find the answer of a quarter of LSC'21 queries (26%) in the first image in the returned ranked list by using only the first hint. Because the result panel in our user interface can show more than 10 retrieved images on a page, users can actually find the answer image in top-1 and top-10 at the same time. This means that E-Myscéal actually helps users solve up to 52% of LSC'21 queries using only one hint, without scrolling through the result panel.

There is a trend that $R@K$ increases as more hints are revealed. However, this is not always the case. For example, R@1 using 4 hints is lower than that using 3 hints with 0.26 and 0.35, respectively. For some queries, the abundance of descriptive details has a negative impact on the performance of the system. This is mainly due to the attempt of the query creator to simulate false memories. For instance, a query from last year's challenge contains the time information of "15th **May** 2015", which was supposed to be "15th **March** 2015". Since there are no data on 15 May 2015, the system does not return any images, making $R@K = 0$. Another explanation could be that the embedding model is not suitable for such descriptions.

To emulate the real challenge, when it is unnecessary to perform more searches once the correct image has been submitted, we modified Table 1 to show a more realistic performance of E-Myscéal in Table 2. In this table, $R@K$ for hint i, denoted as $R@K(i)$, is calculated as $R@K(i) = max(R@K(i), R@K(i-1))$. This means that if the correct image is not in the top-K results in the current hint, we use the $R@K$ of the previous hint. The modified table indicates that users can successfully find the answer of 61% of the LSC'21 queries after all hints are shown without needing to scroll the result panel in our interface many times.

Furthermore, Table 2 shows that there is almost no difference in $R@1$ and $R@100$ when the number of hints used to solve a query increases. Specifically, the difference between these recall figures is 19% if the first hint is used, but it is only 3% when using all 6 hints. This is because by using more hints, the queries become easier to solve and hence the correct images tend to appear first in the result returned by the system.

Table 2: Modified Mean $R@K$ for LSC'21 queries.

Hint	R@1	R@3	R@5	R@10	R@20	R@50	R@100
1	0.26	0.30	0.33	0.38	0.41	0.43	0.45
2	0.43	0.43	0.44	0.45	0.46	0.47	0.49
3	0.49	0.49	0.49	0.50	0.51	0.52	0.52
4	0.52	0.53	0.53	0.54	0.55	0.55	0.56
5	0.56	0.56	0.56	0.57	0.57	0.58	0.58
6	0.58	0.59	0.59	0.60	0.60	0.61	0.61

7 CONCLUSION

In this paper, we describe a new version of Myscéal called E-Myscéal. The new system integrates a new search engine. The embedding technique employed in E-Myscéal is expected to facilitate the retrieval experience of novice users. Compared to Myscéal, the new search mechanism in E-Myscéal does not require users to modify the query many times before entering it into the search bar. Our experiment shows the effectiveness of the new embedding-based search engine in E-Myscéal, which can highly rank the correct answers. Our non-faceted user interface is also updated to remove components that can be confusing to novice users. E-Myscéal, with the simple interface and the new search engine, is expected to help novice users perform well in the LSC'22 challenge.

ACKNOWLEDGMENTS

This publication has emanated from research supported in party by research grants from Science Foundation Ireland under grant numbers SFI/12/RC/2289, SFI/13/RC/2106, 18/CRT/6223 and 18/CRT/6224.

REFERENCES

[1] Kiyoharu Aizawa, Yuto Maruyama, He Li, and Chamin Morikawa. 2013. Food balance estimation by using personal dietary tendencies in a multimedia food log. *IEEE Transactions on multimedia* 15, 8 (2013), 2176–2185.

[2] Naushad Alam, Yvette Graham, and Cathal Gurrin. 2021. Memento: A Prototype Lifelog Search Engine for LSC'21. In *Proceedings of the 4th Annual on Lifelog Search Challenge*. 53–58.

[3] Ahmed Alateeq, Mark Roantree, and Cathal Gurrin. 2021. Voxento 2.0: A Prototype Voice-Controlled Interactive Search Engine for Lifelogs. In *Proceedings of the 4th Annual on Lifelog Search Challenge* (Taipei, Taiwan) *(LSC '21)*. Association for Computing Machinery, New York, NY, USA, 65–70.

[4] Wei-Hong Ang, An-Zi Yen, Tai-Te Chu, Hen-Hsen Huang, and Hsin-Hsi Chen. 2021. LifeConcept: An Interactive Approach for Multimodal Lifelog Retrieval through Concept Recommendation. In *Proceedings of the 4th Annual on Lifelog Search Challenge*. 47–51.

[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] Aaron Duane and Bjorn Þór Jónsson. 2021. ViRMA: Virtual Reality Multimedia Analytics at LSC 2021. In *Proceedings of the 4th Annual on Lifelog Search Challenge*. 29–34.

[8] Olga Gelonch, Mireia Ribera, Núria Codern-Bové, Sílvia Ramos, Maria Quintana, Gloria Chico, Noemí Cerulla, Paula Lafarga, Petia Radeva, and Maite Garolera. 2019. Acceptability of a lifelogging wearable camera in older adults with mild cognitive impairment: a mixed-method study. *BMC geriatrics* 19, 1 (2019), 110.

[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, 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.

[13] 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.

[14] Cathal Gurrin, Björn Þór Jónsson, Klaus Schöffmann, Duc-Tien Dang-Nguyen, Jakub Lokoč, Minh-Triet Tran, Wolfgang Hürst, Luca Rossetto, and Graham Healy. 2022. Introduction to the Fifth Annual Lifelog Search Challenge, LSC'22. In *Proc. International Conference on Multimedia Retrieval (ICMR'22)*. ACM, Newark, NJ.

[15] Silvan Heller, Ralph Gasser, Mahnaz Parian-Scherb, Sanja Popovic, Luca Rossetto, Loris Sauter, Florian Spiess, and Heiko Schuldt. 2021. Interactive Multimodal Lifelog Retrieval with vitrivr at LSC 2021. In *Proceedings of the 4th Annual on Lifelog Search Challenge*. 35–39.

[16] Omar Shahbaz Khan, Aaron Duane, Björn Þór Jónsson, Jan Zahálka, Stevan Rudinac, and Marcel Worring. 2021. Exquisitor at the Lifelog Search Challenge 2021: Relationships Between Semantic Classifiers. In *Proceedings of the 4th Annual on Lifelog Search Challenge*. 3–6.

[17] Basel Kikhia, Andrey Boytsov, Josef Hallberg, Håkan Jonsson, Kåre Synnes, et al. 2014. Structuring and presenting lifelogs based on location data. In *International Symposium on Pervasive Computing Paradigms for Mental Health*. Springer, 133–144.

[18] Emil Knudsen, Thomas Holstein Qvortrup, Omar Shahbaz Khan, and Björn Þór Jónsson. 2021. XQC at the Lifelog Search Challenge 2021: Interactive Learning on a Mobile Device. In *Proceedings of the 4th Annual on Lifelog Search Challenge*. 89–93.

[19] Xirong Li, Chaoxi Xu, Gang Yang, Zhineng Chen, and Jianfeng Dong. 2019. W2vv++ fully deep learning for ad-hoc video search. In *Proceedings of the 27th ACM International Conference on Multimedia*. 1786–1794.

[20] Jakub Lokoč, František Mejzlík, Patrik Veselý, and Tomáš Souček. 2021. Enhanced SOMHunter for Known-item Search in Lifelog Data. In *Proceedings of the 4th Annual on Lifelog Search Challenge*. 71–73.

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

[23] Manh-Duy Nguyen, Binh T Nguyen, and Cathal Gurrin. 2021. A Deep Local and Global Scene-Graph Matching for Image-Text Retrieval. *arXiv preprint arXiv:2106.02400* (2021).

[24] Manh-Duy Nguyen, Binh T Nguyen, and Cathal Gurrin. 2021. Graph-Based Indexing and Retrieval of Lifelog Data. In *International Conference on Multimedia Modeling*. Springer, 256–267.

[25] Thao-Nhu Nguyen, Tu-Khiem Le, Van-Tu Ninh, Minh-Triet Tran, Nguyen Thanh Binh, Graham Healy, Annalina Caputo, and Cathal Gurrin. 2021. LifeSeeker 3.0: An Interactive Lifelog Search Engine for LSC'21. In *Proceedings of the 4th Annual on Lifelog Search Challenge*. 41–46.

[26] Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, et al. 2021. Learning transferable visual models from natural language supervision. In *International Conference on Machine Learning*. PMLR, 8748–8763.

[27] Luca Rossetto, Matthias Baumgartner, Narges Ashena, Florian Ruosch, Romana Pernischová, and Abraham Bernstein. 2020. Lifegraph: a knowledge graph for lifelogs. In *Proceedings of the Third Annual Workshop on Lifelog Search Challenge*. 13–17.

[28] Luca Rossetto, Matthias Baumgartner, Ralph Gasser, Lucien Heitz, Ruijie Wang, and Abraham Bernstein. 2021. Exploring Graph-querying approaches in LifeGraph. In *Proceedings of the 4th Annual on Lifelog Search Challenge*. 7–10.

[29] 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.

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

[31] Erich Schubert, Jörg Sander, Martin Ester, Hans Peter Kriegel, and Xiaowei Xu. 2017. DBSCAN revisited, revisited: why and how you should (still) use DBSCAN. *ACM Transactions on Database Systems (TODS)* 42, 3 (2017), 1–21.

[32] L. Signal, M. Smith, M. Barr, J. Stanley, T. Chambers, Jiang Zhou, A. Duane, G. Jenkin, A. Pearson, C. Gurrin, A. Smeaton, J. Hoek, and C. Ni Mhurchu. 2017. Kids'Cam: An Objective Methodology to Study the World in Which Children Live. *American Journal of Preventive Medicine* 53 3 (2017), e89–e95.

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

[34] Florian Spiess, Ralph Gasser, Silvan Heller, Luca Rossetto, Loris Sauter, Milan van Zanten, and Heiko Schuldt. 2021. Exploring Intuitive Lifelog Retrieval and Interaction Modes in Virtual Reality with vitrivr-VR. In *Proceedings of the 4th Annual on Lifelog Search Challenge*. 17–22.

[35] Ly-Duyen Tran, Manh-Duy Nguyen, Nguyen Thanh Binh, Hyowon Lee, and Cathal Gurrin. 2020. Myscéal: An Experimental Interactive Lifelog Retrieval System for LSC'20. In *Proceedings of the Third Annual Workshop on Lifelog Search Challenge*. 23–28.

[36] Ly-Duyen Tran, Manh-Duy Nguyen, Nguyen Thanh Binh, Hyowon Lee, and Cathal Gurrin. 2021. Myscéal 2.0: A Revised Experimental Interactive Lifelog Retrieval System for LSC'21. In *Proceedings of the 4th Annual on Lifelog Search Challenge*. 11–16.

[37] Hoang-Phuc Trang-Trung, Thanh-Cong Le, Mai-Khiem Tran, Van-Tu Ninh, Tu-Khiem Le, Cathal Gurrin, and Minh-Triet Tran. 2021. Flexible Interactive Retrieval SysTem 2.0 for Visual Lifelog Exploration at LSC 2021. In *Proceedings of the 4th Annual on Lifelog Search Challenge*. 81–87.