Myscéal: A Deeper Analysis of an Interactive Lifelog Search Engine

Ly-Duyen Tran[1†], Manh-Duy Nguyen[1†], Binh T. Nguyen[2,3*] and Liting Zhou[1]

[1]Dublin City University, Ireland.
[2]AISIA Research Lab, Vietnam.
[3]VNU HCM - University of Science, Vietnam.

*Corresponding author(s). E-mail(s): ngtbinh@hcmus.edu.vn;
Contributing authors: ly.tran2@mail.dcu.ie;
manh.nguyen5@mail.dcu.ie; zhou.liting2@mail.dcu.ie;
[†]These authors contributed equally to this work.

Abstract

There is a growing number of lifelogging retrieval systems that have been introduced in several lifelogging workshops and events. Across all systems at the LSC, which is an annual international challenge on lifelogging retrieval, our Myscéal is currently considered as the state-of-the-art. In this paper, we describe the system in detail and show how it has been upgraded through time since firstly introduced in 2020. In addition, we analyse Myscéal performance not only in the three lifelog retrieval competitions it participated in but also with additional user experiments. The result shows that the fast searching time of Myscéal is the system's most important feature that helps it get some significant advantages in competitions. On the other hand, the findings from user experiments indicate that Myscéal still needs some improvements for novice users who are unfamiliar with how to interact with the system. Moreover, the user study plays a vital role in the development of Myscéal as many updates of this system came from the feedback of the participating users. We also demonstrate the efficacy of Myscéal as a lifelog retrieval system to help the lifeloggers, who capture their daily life in images, recall memorable moments in their massive lifelog archives.

Keywords: lifelog, interactive retrieval system, human computer interaction

1 Introduction

A simple and portable camera or smartphone can capture hundreds of images daily. People are using it to keep track of every moment happening in their daily life, which is termed lifelogging [1]. This results in a large lifelog image dataset of thousands of images per day. Lifelogging raises a need to find an efficient way of retrieving images within that massive archive whenever they want to recall some specific moment in the past. To address this issue, many fora about developing a retrieval tool that can assist them have been introduced; for instance, the Lifelog Search Challenge [2], ImageCLEFlifelog [3], and NTCIR-Lifelog [4]. These collaborative benchmarking sessions not only facilitate sharing a lifelog dataset but also propose a research challenge reflecting the retrieving use case. The challenge requires participating retrieval systems to find relevant images of the moment described by a given semantic query. Lifelog retrieval systems attending to these challenges should produce an accurate result after an appropriate time. Myscéal [5] is built to correspond to these criteria when it has a powerful search engine and a straightforward interface that allows users to operate efficiently. These features can help Myscéal provide great results without requiring many interactions. The system was introduced in its first version in 2020 and followed the standard approach of typical lifelog retrieval systems in the preceding years: annotating lifelog images with the visual objects appearing within them. Based on those annotations, a retrieval system can compare with keywords in the given query to return relevant images using similarity metrics. Our Myscéal, distinctly from all systems, is implemented with a novel feature that considers the area of the visual objects in addition to its semantic labels as others. This state-of-the-art lifelog retrieval application comes with a clear user interface design aimed at novice users unfamiliar with the system. We utilise the space in the interface as much as possible to show the returned images by removing most of the faceted filtering area and combining them into one single search box. The clean and simple interface is expected to help users not be confused with many incidental areas but only focus on the main screen showing the retrieved images.

In this work, our contribution is threefold. Firstly, we describe how Myscéal works in detail from the initial annotation step to the indexing stage in the database with examples. Since Myscéal has been updated with several features for each version, we also summarise those features and show how it has improved since its first release in 2020. Secondly, the performance of Myscéal in three lifelog search competitions is analysed to show which factors contributed the most to this state-of-the-art lifelog retrieval system to achieve significant results in those challenges. Finally, we recap a Myscéal novice user study to show how impactful this experiment is on the updates of Myscéal after each iteration. We also examine the user experiment on the lifelogger, the true target of any lifelog retrieval system, to show the efficacy of Myscéal in helping them retrieve images from their own lifelog dataset.

2 Related Works

The most common workflow implemented by many current lifelog retrieval systems has been adopted from the LeMore system [6] which can be considered as one of the first interactive lifelog retrieval systems. The pipeline behind the LeMore was to enrich images by annotating visual object labels detected in images and matching them with a query description. Many systems following this method have won some lifelog retrieval challenges [7–9]. This result has indicated the effectiveness of this conventional method. Furthermore, the annotated information also can be used as a filter mechanism to enhance the retrieval results produced with a visual-based input sketched by users [10, 11]. Another approach is to embed both lifelog images and textual queries into the same vector space where their relevance can be measured by a similarity metric such as cosine distance. Some systems that applied this method have achieved remarkable performance in some lifelog retrieval challenges [12–14]. Additionally, knowledge graphs can be employed to extend annotated visual objects to get their synonyms for better retrieval [15] or to capture interaction information in the images [16]. In addition to retrieval methods, the user interface also contributes to the performance of a lifelog retrieval system. Searchers can retrieve images quickly by using a system built in the virtual environment which can help them perform the retrieval based on their gesture [8]. Due to a potentially large number of results, it is a challenge to display all of them in a meaningful way. The self-organizing maps (SOM), a comprehensible display technique for many images, have been applied and improved the search time [17, 18].

Myscéal viewed this as a document retrieval problem by indexing textual annotations of images and matching them with textual queries. Although we follow existing systems by annotating images with visual objects, we further extend them when considering the area of objects within images beside their semantic labels. Therefore, we introduce a novel scoring feature that emphasises bigger objects over small ones. Regarding the user interface, Myscéal is designed with a straightforward scheme. We remove all faceted filters and buttons and replace them by combining them into one simple search box. Searchers can simply input keywords to retrieve or filter within the same single box. Additionally, Myscéal combines a sequence of similar temporal images into one image instead of showing all images at once to the user. This is to avoid confusion and maximise the utility of the interface's display area.

3 Comparative Benchmarking for Lifelog Retrieval

In this section, we introduce the two relevant competitions about the lifelog retrieval field in which Myscéal [19] already participated.

3.1 Lifelog Search Challenge

Although not being the first challenge about lifelogging retrieval, the **Lifelog Search Challenge** [2] (LSC) is the pioneer in interactive benchmarking in regard to this lifelog research field, with the first edition in 2018. The LSC is organised to create a comparative evaluation of interactive lifelog retrieval systems in an open, real-time, and metrics-driven evaluation in which all systems need to compete at the same time to solve the same set of queries. From here, we will refer these given queries as *tasks* to differentiate them from the actual *queries* entered by the users while performing searches. The lifelog dataset used in the LSC consists of anonymised lifelog images captured by only one lifelogger during one month (LSC'18 and LSC'19) or 3 months (LSC'20 and LSC'21). The images were synchronised with their timestamp and locations, indicating when and where they were taken. Along with the images, a set of concept annotations of each image were released. In addition, the lifelogger's biometric data and physical activity such as walking or driving were also recorded. This international content retrieval competition requires participating lifelog retrieval systems to seek images relevant to given tasks created by the lifelogger who is the dataset's owner. Each task indicates a specific memorable moment or event that occurred in the lifelogger's daily life. Because the LSC is a real-time and interactive challenge, each task is not shown to teams entirely at once but gradually by providing additional clues every 30 seconds. A task will be revealed partially through 5 iterations, starting with a vague piece of information and getting more detailed with every iteration, gradually making it easier for participants to solve the tasks. Some example LSC tasks can be found in Table 3. Regarding evaluation metrics, the LSC considers the accuracy and search time, meaning that the retrieval systems attending the challenge need to find the correct answer in a short amount of time with the least wrong submissions. Furthermore, the LSC encourages participating teams to develop easily operable lifelog retrieval systems. Therefore, the challenge has two different sessions for expert and novice users, and each session will have a similar evaluation. The top-performing systems need to retrieve the results correctly and quickly and have a user-friendly user interface that can help novice users interact with it with ease. Nevertheless, the two recent LSC'20 and LSC'21 only had expert sessions due to the COVID pandemic.

3.2 ImageCLEFlifelog

ImageCLEFlifelog [3] is another competition for lifelog and was introduced in 2017, one year earlier than the LSC. In addition to the ordinary lifelogs retrieval research, this challenge has another task in the lifelog summarisation field. This task encourages participants to explore the possibilities that come with lifelog data by analysing the data in a deeper manner. However, since Myscéal is a lifelog retrieval system, we only focus on the retrieval task of this competition. Although both competitions are about lifelogging retrieval, ImageCLEFlifelog is different from the LSC that the former is organised in the off-line setting. All semantic tasks used in the ImageCLEFlifelog will be given to participants at once. The challenge requires retrieval systems to not only find one specific image but also focus on seeking all images relevant to a given task. Because of that configuration, this challenge considers the harmonic mean of the recall and the precision of the retrieved images as the evaluation metrics and does not focus on the retrieval time as the LSC does. The dataset used in ImageCLEFlifelog is moderately similar to that of the LSC. In particular, the dataset comes with thousands of lifelogging images and their automatically extracted visual concepts. Some other semantic contents are also provided, such as locations or activities. Music listening history and biometrics information, such as heart rate, calories burned, or steps, are also recorded.

4 Myscéal

The pipeline of our system is illustrated in Figure 1. Visual concepts extracted from the images and activity descriptors, GPS coordinates, and semantic locations are used to create an inverted index in the ElasticSearch engine. User interactions are transformed into ElasticSearch queries. The user interface is designed to present the results with extra information in a straightforward manner, which allows the user to quickly select and submit the targeted image to the evaluation server. In this section, we will describe Myscéal by explaining each component in Figure 1, consisting of *processing*, *search* and *user interface* components.

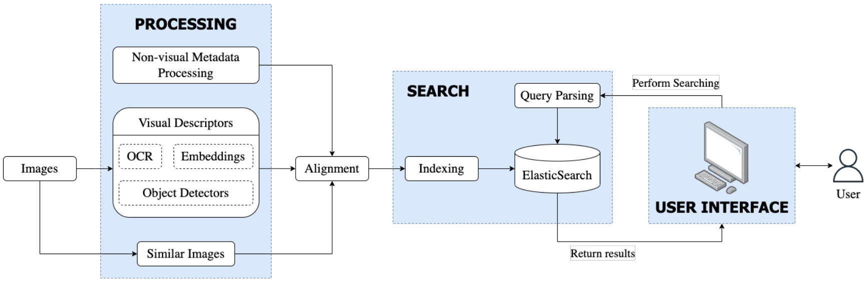

Fig. 1 Pipeline of Myscéal which contains three main components: Processing, Search, and User Interface.

4.1 Processing components

4.1.1 GPS clustering

There are 421 different named places in the metadata of the whole dataset. However, depending on the locations mentioned in the query, we only choose a subset to visualise on the map.

4.1.2 Visual descriptors

Following the standard approach of enriching visual concepts of many systems in previous years, we decided to utilise additional features from existing state-of-the-art computer vision models as follows:

- Semantic labels from DeepLabv3+ [20], an image semantic segmentation model. This process labels each pixel in an image with a corresponding visual concept, resulting in a segmentation map.
- Object detection and image captioning concepts from Microsoft Computer Vision. API [1],
- Optical Character Recognition (OCR) results from Google Cloud Vision API [2].
- Material and colour concepts from Bottom-up attention model [21] trained on Visual Genome Dataset [22].

Furthermore, Myscéal exploits the *area* of each visual object in an image, or in other words, the pixel count in the semantic segmentation result and the bounding box area in the object detection result[3]. The idea behind this is the assumption that bigger objects play a more important role in an image. This information is used to in a scoring scheme called **aTFIDF**[5].

For a given image, if we denote the set of images from the LSC'20 dataset by I, the collection of possible object keywords 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:

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

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.

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

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

[1] https://azure.microsoft.com/en-us/services/cognitive-services/computer-vision/
[2] https://cloud.google.com/vision
[3] Here, we also consider an OCR text as a visual object with its bounding box.

4.1.3 Non-visual metadata processing

The lifelog dataset used in LSC'20 and LSC'21 consists of images and associated non-visual metadata, namely, GPS coordinates, time in UTC, as well as (misaligned) local time and activity recognition from biometrics data (walking, transport, etc.). Using the provided GPS coordinates, we used Google Maps API's Reverse Geocoding service [4] to obtain addresses. This allows filtering using the names of states, countries, and unannotated semantic locations. All location-related texts are normalised to ASCII. Regarding time data, we converted UTC time to local time based on the time zone detected from the GPS coordinates and realigned the images based on semantic locations. The days of the week (Monday, Tuesday, etc.) are also extracted from time information. Myscéal does not use other data like music or heart rate, as this information did not contribute much to the previous LSC events.

4.1.4 Temporal units

To describe how Myscéal works, we define a temporal hierarchy of events consisting of three units: **image**, **scene**, and **event**. The smallest temporal unit is an **image**, which is an atomic unit of a lifelog. We consider a **scene** to be the combination of one or many subsequent similar *images*. An example is when the lifelogger is working at a desk, and his surroundings remain practically unchanged. An **event** consists of one or multiple consecutive *scenes*, whose boundaries are indicated either by a change of contexts, such as location and activity or by a significant time gap. These units are used throughout indexing, searching, and user's interacting.

We segment lifelog into *events* using the provided location semantic names and activities. To define the segmentation boundary of *scenes*, we assign each image three feature vectors including VGG16 feature [23], Word2Vec feature [24], and SIFT feature [25, 26]. We compare adjacent images by calculating three cosine distances and building a Naive-Bayes classifier to determine scene boundaries using these distances.

4.2 Search components

4.2.1 ElasticSearch indexing

We employ an off-the-shelf search engine called ElasticSearch. ElasticSearch, an open-source search and analytic engine, supports searching with varied data types. The lifelog index is created as a collection JSON-like documents with the properties shown in Table 1 in the `scene` index. This database provides filterable information, as seen in Table 1, and is used as a way to narrow down the search space before more complex calculations are applied to each image. The main `image` index is created with more fine-grain properties, as seen in Table 2.

[4] https://developers.google.com/maps/documentation/geocoding/requests-reverse-geocoding

Table 1 Elasticsearch document for each scene.

	Explanation	ES Format	Examples
`images`	the list of image IDs	keyword	`20160810_071508_000`, `20160810_071421_000`
`begin_time`	local time	date	2016/08/10 08:12:00+00
`end_time`	local time	date	2016/08/10 08:12:00+00
`desc`	the list of visual concepts with equal importance	keyword	station, red wall
`weekday`	the day of the week	keyword	monday
`location`	provided semantic name of the location	keyword	home, angelica's cafe
`address`	reverse geocoding result	text	whitehall, dublin, ireland
`gps`	GPS coordinates	geo_point	53.3858, -6.2607
`activity`	provided activity recognition	keyword	transport

Table 2 Properties of Elasticsearch document for each image.

	Explanation	ES Format	Examples
`image_id`	the image ID	keyword	`20160810_071508_000`
`time`	local time	date	2016/08/10 08:12:00+00
`atfidf_s`	aTFIDF feature from semantic segmentation	rank_features	`{"wall": 1.35, "person": 6.79}`
`atfidf_o`	aTFIDF feature from object detection	rank_features	`{"wall": 1.35, "person": 6.79}`
`atfidf_ocr`	OCR feature	rank_features	`{"online": 16.892, "book": 18.00}`

4.2.2 Query processing

To reduce the number of actions a user interacts with the system, we decided not to use a faceted interface to show filters in different metadata such as date, time, and location. Using ad hoc regular expression patterns, we map the textual query into corresponding fields. Moreover, after getting a list of visually descriptive words from the query, we use Word2vec[24] and WordNet[27] to map the words into a specific and limited set of keywords provided by object detectors and semantic segmentation engines. For example, *tea* might imply the presence of a *mug*, or a *teapot*. This process transforms every concept into a list of keywords with the corresponding relevance scores.

4.2.3 Primary search mechanism

For a processed query, time-related information (day of the week, date, month, year, or time of the day), locations of large areas(cities and countries), and activities are used as filters. Semantic locations are used as GPS filters based on the result of GPS clustering instead of pure text matching, based on the assumption of incomplete annotation. Myscéal's backend first applies these filters to the `scene` index. Then, the system scores the remaining scenes as follows:

For a list of visual concepts, $q = [q_0, q_1, ..., q_m]$, where each concept q_i is expanded in to a list of keywords $[o_{i,0}, o_{i,1}, ..., o_{i,n}], o_{i,j} \in O$ with the relevance scores of $[r_{i,0}, r_{i,1}, ..., r_{i,n}]$, the final score of each scene can be formulated as:

$$S_{\text{scene}} = \sum_{i=0}^{m} \max_{\{j | o_{i,j} \in \text{scene}\}} r_{i,j}$$

After that, we have narrowed down the search space to the top N scenes. Using the images belonging to these scenes, we combine the scores on the three fields `atfidf_s`, `atfidf_o`, and `atfidf_ocr` in Table 2. Each score is calculated as follows:

$$S_{\text{image}} = \sum_{i=0}^{m} \max_{j} (r_{i,j} * \text{aTFIDF}_{\text{image}}(o_{i,j}))$$

4.2.4 Complementary search mechanism

Aside from using a single textual input as search query, Myscéal offers other means of search.

Temporal search After analysing the tasks from previous challenges, we chose to support users to search for multiple time-related events. The system at first searches for one of the events and uses the resulting time information as a conditional input to search for the second event. Then, the process is repeated in the reverse direction to enhance the recall performance. All results will be grouped at the last step and ranked based on their total score.

GPS search The search query can also be extended with a location filter using a bounding box of GPS coordinates. In the case of multiple temporal queries, the filter is only applied on the main query.

Visual similarity search Visual similarity can help the user find visually similar images to any given image. The similarity scores are calculated using the cosine distance of a concatenation of the SIFT [25, 26, 28] and VGG16 [23] features.

4.3 User Interface components

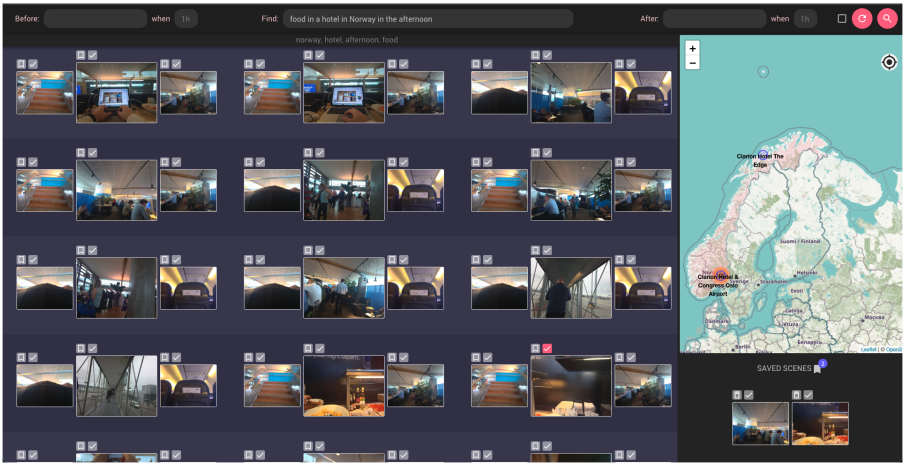

Fig. 2 User interface of Myscéal

Due to the expectation of novice users' involvement, the UI was designed with two main principles: to minimise different steps in the search process and allow back-end functionality to be fully used. The UI can be seen in Figure 2.

4.3.1 Search boxes

The temporal relationships between different query clues can be specified using the three boxes at the top. The primary query is placed in the middle and can be entered quickly. Furthermore, the time conditions (in hours) of "before" and "after" queries can be specified in the small boxes.

4.3.2 Query suggestion

Since the second version of Myscéal, we have exposed the query expansion to help the user adjust the query accordingly. The second version [29] allows the user to modify the relevance score of each visual concept or remove the concept altogether. However, this feature is removed in the final version. In the LSC'21 version, we show the interpretation of the query under the search boxes and highlight if a word does not appear in the indexed database, prompting the user to double-check that word and select another option in the suggested list if necessary.

4.3.3 Search results display

A large proportion of the screen is dedicated to displaying the search result. Each resulting entry is arranged corresponding to the temporal relationship, which is the main event to be searched for in the middle. As we segmented lifelog data into scenes, each thumbnail here represents a scene consisting of multiple images. The user can access all images belonging to a scene by clicking on the thumbnail (as illustrated in Figure 3).

4.3.4 Map

Another critical part of the UI is a map, as seen in Figure 2. By default, it shows the locations of the retrieval result. However, it also supports GPS search in an area by allowing the user to draw a rectangle on the map specifying the intended area. Location names, including semantic locations such as home or workplace related to the query, are also shown on an overlay layer of the map.

4.3.5 Visual Similarity and Event View

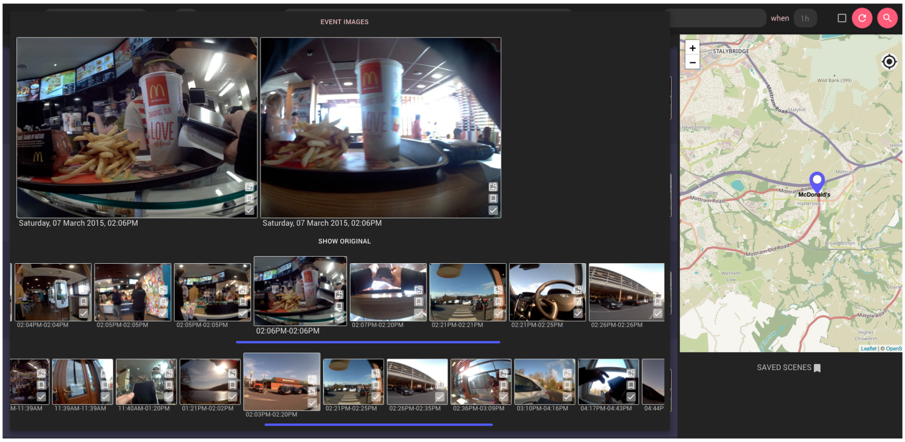

Fig. 3 Events view window

Figure 3 shows an pop-up panel showing lifelog images in a temporal context. This view can be accessed by clicking on any image shown in the main interface in Figure 2. The event view presents the hierarchy of the three temporal units mentioned in Section 4.1.4, allowing the user to browse the images from that day at two different paces.

Moreover, a user can search for similar images by clicking on a small visual similarity icon at the bottom of each photo here.

5 Myscéal performance

In this section, we highlight some results that Myscéal achieved and how the system performed in the challenges mentioned in Section 3.

5.1 LSC'20

LSC'20 is the third edition of the LSC series, and also the first lifelog retrieval challenge that Myscéal participated in. Despite being the new retrieval system in the competition, Myscéal obtained the highest overall score among 14 participants and achieved the first place in LSC'20. Figure 4 illustrates the precision and recall of all teams in the challenge, in which the order of teams indicates their final ranking. The precision is defined as the portion of correct submissions out of the total submission in the competition. Meanwhile, recall is the percentage of tasks that a team managed to solve successfully. We can observe in Figure 4 that Myscéal and SOMHunter [17] got the highest recall compared to others at 87.5% meaning that both systems managed to solve 21/24 tasks. Moreover, Myscéal also had the highest precision metric at 84%, which indicates that the system only submitted a few wrong answers with four incorrect submissions out of 25 submissions in total during the competition.

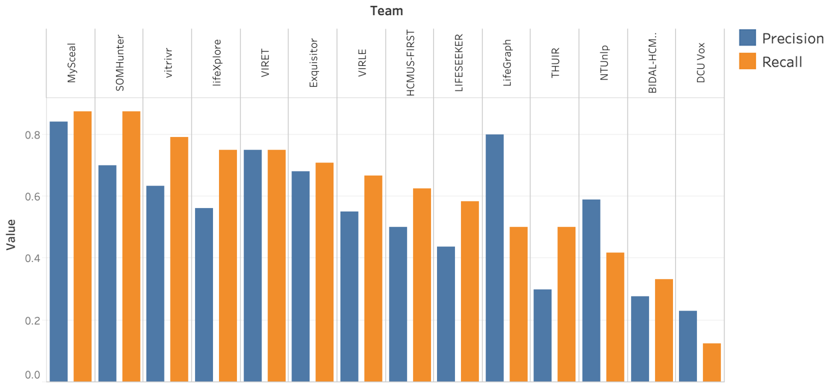

Fig. 4 Precision and Recall of each team in LSC'20.

The LSC also evaluates systems by taking into account the speed of submission (faster is better). Figure 5 depicts the retrieval speed of participants in LSC'20. Myscéal was in the top three quickest systems to return the correct result 13 times, which was the highest across all participants and significantly higher than the second system. This means that half of the time in the competition (13/24) Myscéal found the correct answers in the top-3 fastest systems. This search time criterion is one of the key factors helping Myscéal obtain the first place within the LSC'20.

Due to the COVID pandemic, LSC'20 was the first time in this series organised virtually. Therefore, the challenge only had a session for experts and did not have a session for novice users as it had facilitated in previous years. We conducted a user study using Myscéal with eight novice participants to have more insights into the system's performance when used by novice users. Furthermore, their feedback is a valuable resource from which we can build a

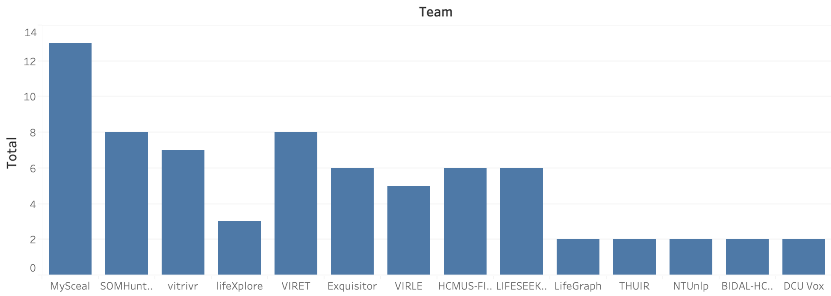

Fig. 5 Number of times each team was in the top-3 quickest team to return the correct results.

better system for LSC'21 with additional features. The setting of the experiments is replicated from the live campaign regarding the time limit and the order in which the clues are shown. However, only five tasks were chosen from the official list of tasks used in LSC'20 to keep the experiment sessions short. These tasks are illustrated in Table 3. They were arranged in an ascending level of difficulty based on our experience in the official competition. Before doing the experiment, all novice users were briefly instructed to learn how to operate Myscéal by trying to solve 3 sample tasks in a pre-experiment session. These three tasks were also selected from the query bank of the LSC'20.

Table 3 List of tasks used in our novice user study. The tasks were chosen from the query bank used in LSC'20. The symbol '/' separates clues that are gradually revealed to searchers.

Task	Clues
1	Taking a photograph of an A380 airplane/ in Germany/ before boarding a flight/ in the late afternoon/ in 2015/ on the 19th March.
2	It was the best cake I had in years,/ in an antiques store./ I was alone drinking tea and eating cake./ I think I finished all the cake in 3 minutes./ It was in the UK/ on a Saturday morning.
3	I was having beer after a long day of meetings./ It was a 'corona extra' beer in a bottle./ I remember the room was dark./ I was relaxing in a hotel lobby bar./ I don't remember there was anyone else there./ It was in May 2018, in Wuhan.
4	Passing by a clocktower while running/ in a park near my home./ It was in the early morning, around dawn./ I drove to the park/ and I drove home again afterwards./ It was a Saturday morning in February.
5	Four red figures,/ maybe they are aliens./ It looked like a painting of aliens./ There were walking on the desert./ There was a big red wall behind the painting./ And a TV, I think there was also a TV there./ I was having tea and sandwiches in March 2015.

The score of each participant is shown in Table 4, which was calculated using the same formula of the live campaign. The score of Myscéal is the score gained by expert users in the actual LSC'20. As can be observed in Table 4, all users (including experts) fail to find the correct answer to the last task. The first two tasks were easier to solve compared to others when all searchers managed to find the relevant images. As we ordered tasks based on their difficulty, task 1 was the task that users obtained the highest score across all tasks. This is because of the keyword "*airplane*" when there was only a small amount of images containing airplanes within it. Meanwhile, the valuable clue of "*UK*" location, which helped users solve the second task, was only presented in the last 60 seconds, giving them a considerably low score. It was also true for the third task, but there were many images with bottles taken in Wuhan, making users unable to find the correct answer. In contrast, the location "*home*" in the second hint of the fourth task allowed users to have enough time to find the right images. Six novice users solved three tasks, and two users (U6 and U8) got two correct results. This indicates that novice users could use the system without significant issues since even the expert could only find four correct answers. None of the eight novice users could solve the last task, which is as expected since the expert could not make it either in the official competition. This was because Myscéal in LSC'20 could not detect the color in images, and the word "*aliens*" represented a significant semantic and lexical gap between the information need and the dataset. There was still a big gap in the performance between novice and expert users when the average score of novice users was 188.97, which is just a half of the score obtained by the expert user at 339.03.

Table 4 Experiment score of 8 novice users compared to Myscéal team's official score in LSC'20.

Task	**Myscéal**	**U1**	**U2**	**U3**	**U4**	**U5**	**U6**	**U7**	**U8**
1	94.86	92.5	94.86	85	86.81	98.19	87.92	90.69	82.36
2	78.06	54.86	50.69	68.61	47.08	59.03	57.92	43.33	59.58
3	87.5	53.47	0	0	53.33	0	0	0	0
4	78.61	0	77.5	51.11	0	52.5	0	64.44	0
5	0	0	0	0	0	0	0	0	0
Total	339.03	200.83	223.06	204.72	187.22	209.72	145.83	198.47	141.94

The most significant issue of Myscéal used in the LSC'20 challenge was that this system did not implement colour detection. Therefore, none of the users managed to find the answer to Task 5 in Table 3 where the clue about *"red wall"* was the most informative hint. The same was true for the OCR clues. On the other hand, Task 3 in Table 3 could be easily solved if Myscéal had the feature of searching for text to find *"corona extra"*. Additionally, we found that users tended not to use the map area in the UI although the clues about location contained useful information. We overcome these problems by adding some major updates to Myscéal to participate in LSC'21 [19] and LSC'22 [30]. We included the colour detector and the OCR in the annotation processing component. Furthermore, we also enlarged the map area in the UI to encourage users to utilise this unique feature.

5.2 ImageCLEFlifelog'20

Myscéal is developed to match the evaluating criteria of the LSC, which requires a system to find a single specific image that is relevant to a semantic query as quickly as possible with the least number of wrong submissions. However, ImageCLEFlifelog is different as this competition is a more conventional non-real-time retrieval challenge. It requires participating systems to find as many relevant images as possible and does not take the retrieval time into account. We slightly modified Myscéal from LSC'20 by adding an event row at the bottom of the UI as shown in Figure 3. This feature was expected to help users scroll faster to find all relevant images. Moreover, we added a small feature that could help users adjust the scores of input keywords to revise the results [29]. Despite not originally matching the challenge's objectives, Myscéal obtained the third place in ImageCLEFlifelog'20 [29].

This competition is also a good opportunity to evaluate Myscéal performance in different use cases, and we conducted user experiments with three users: an expert, a novice user, and the data owner (the lifelogger). It is essential to point out that ImageCLEFlifelog is a suitable challenge rather than the LSC to include the lifelogger as a user without worrying about their prior knowledge about the dataset. This is because the LSC only needs a lifelog image indicating a specific moment to solve a query, although there are maybe many of them that match the query. This makes it easy for the lifelogger who owns the data to solve, as they can recall the most recent event relevant to the query. However, ImageCLEFlifelog requires a searcher to retrieve all images instead of one. This means that the lifelogger needs to remember all events, which is more difficult for them to solve the query if they use only their memory without using any lifelog retrieval system. For example, to get the maximum points for the query "*Find the moment when the lifelogger was getting a bus to their office*" in ImageCLEFlifelog, the searcher has to find all images belonging to the relevant moments which might happen many times in different days.

ImageCLEFlifelog'20 contained 10 queries as tasks to be solved[5]. For each task, searchers need to find the top 100 images belonging to all relevant moments matched with the corresponding query. In our experiment, each of the three users had a total of 5 minutes to solve a task, reading time not included. The data owner and the novice user were quickly instructed to learn how to use Myscéal prior to the experiments. We used the same evaluation metric in ImageCLEFlifelog'20, which is the F1@10 score. In order to get the highest F1@10 score (which is 1) of a task, the top-10 images of the result should belong to all events described by the query.

Table 5 shows the score of all users in the experiments. We can see that the expert, who had the advantage of knowing the system and being familiar with most of the dataset, achieved the highest score. Despite having no experience

[5] https://www.imageclef.org/system/files/ImageCLEF2020-test-topics.pdf

Table 5 $F1@10$ scores of 3 users (U1: Lifelogger, U2: Expert, U3: Novice). The symbol '−' indicates that the user was unable to find the answer for that task. The numbers with $*$ are the highest number in that topic.

Task	U1 (lifelogger)	U2 (expert)	U3 (novice)
1	0.58	1*	0.67
2	0.72*	0.22	-
3	1*	0.57	1*
4	0.31*	0.22	-
5	0.68*	0.68*	-
6	0.25	0.5*	0.25
7	0.69	0.89*	0.69
8	0.75	1*	-
9	0.8*	0.73	0.77
10	0.5	0.5	0.5
Overall	0.63*	0.63*	.39

with the system, the data owner obtained comparable scores in most tasks and got the same overall score. The average $F1@10$ score of the novice user was lower than that of the others due to the fact that this user was unsuccessful in solving nearly half of the tasks in the challenge. Additionally, although having knowledge of the dataset, the lifelogger got three tasks with the highest F1@10 score, which was lower than that of the expert at 4.

Although the lifelogger and the expert user successfully solved all tasks, the novice user only found the answer for half of them. This opened the question of how effective the Myscéal interface was. Having another scrolling bar for event viewing could confuse novice users. We observed that both the lifelogger and the novice user rarely used this feature. Furthermore, the implemented keyword scoring adjustment feature was not helpful as expected when both users completely ignored this function. It did not contribute much to the result of the expert user when the revised result after modifying the weights was not relevant to the queries. Therefore, we decide to remove this feature from Myscéal in LSC'21 [19] and LSC'22 [30]. We also make some changes to the interface for Myscéal in LSC'22 to make it simpler for novice users [30].

5.3 LSC'21

Myscéal participated in LSC'21 with additional features and updates in the user interface, which were based on comments and feedback from novice users in our user study described in Section 5.1 and 5.2. Like the previous iteration, LSC'21 was a virtual competition, hence could not have a novice session. There were 23 tasks used in LSC'21, roughly similar to LSC'20 at 24 tasks. The number of participating systems increased from 14 to 17 in LSC'21. The other settings of the competition remained the same as in its previous campaign.

Table 6 Summary of LSC'21 result of top-6 systems. The numbers in bold are the highest number among the top-6 systems.

	MySceal	SomHunter+	LifeSeeker	Voxento	CVHunter	Memento
Solved tasks	19	19	**20**	18	15	16
Wrong submission	4	9	6	**3**	8	11
Precision (%)	82.61	67.85	76.92	**85.71**	65.21	59.25
Recall (%)	82.61	82.61	**86.95**	78.26	65.21	69.56
Submitted in Top-3	**12**	**12**	9	11	5	9
Overall Score	**100**	97.6	97	91.4	77.3	77.2

Myscéal obtained the first place in LSC'21 as it did in the previous year. However, LSC'21 witnessed competitive performance between teams when differences in the scores of the top-3 systems were minuscule. Summary scores of the top-6 systems in LSC'21 are illustrated in Table 6 in which precision and recall are defined as discussed in Section 5.1.

The Submitted in Top-3 indicates how fast the systems performed. This is the number of times that a system manages to be in the top-3 speediest systems to find the correct answer.

The overall score is the normalisation of the total score awarded by solving the tasks in the competition. This normalised score is the main metric used to rank the systems in the LSC. Table 6 shows that there is a negligible gap in the scores of Myscéal, SomHunter+ [18], and Lifeseeker [31]. Although Myscéal attained the best overall score, Lifeseeker was the team that solved the most tasks (20/23) in the challenge and got the highest recall at 86.95%. Regarding precision, Myscéal was not the best in this metric either when Voxento only had three wrong submissions, making it the highest precision at 85.71%. Myscéal, with 19/23 successfully solved tasks and three incorrect submissions, had the same precision as with recall at 82.61% for both metrics. It is interesting that Myscéal was not the system that solved the most number of tasks nor had the least wrong submission, yet managed to win the competition. This is because Myscéal was one of the fastest systems that could find the correct answer compared to others. As shown in Table 6, Myscéal and SomHunter+ were the two systems that had the highest times submitting the correct answer in the top-3 fastest systems in the competition with 12 times.

One of the essential features of Myscéal in LSC'21 was OCR and colour detection. Half of the tasks in LSC'21 included the OCR clues from which participating systems could easily find the correct answer. Furthermore, we used the similar image search function many times in LSC'21 to find the relevant result from the initial result. Therefore, we continuously integrate these features for LSC'22, but there are some changes in the UI where we make the similar images panel easier for users to access and explore. Another critical update is that we will change the approach of Myscéal for LSC'22 [30]. LSC'21 witnessed the effectiveness of embedding techniques when SOMHunter+, Voxento, and Memento quickly solved tasks describing activities that were difficult for Myscéal to find answers. Therefore, instead of using keywords as in previous versions, we change Myscéal to E-Myscéal [30] which applies an embedding approach to participate in LSC'22.

6 Discussion

Myscéal is originally developed for the LSC competition, which is to quickly find a single image that is relevant to a semantic query. The result of Myscéal in LSC'20 and LSC'21 has shown the system's efficacy with the powerful search engine and the straightforward user interface. Moreover, both compartments helped Myscéal to win the two most recent iterations of the LSC competition by solving most of the tasks faster than other teams. Nevertheless, while Myscéal surpassed other systems in terms of precision, recall, and search time by a large margin in LSC'20, Myscéal achieved the first place in LSC'21 with a tiny difference to other teams when this system did not perform significantly better in any metrics. LSC'21 witnessed a rise in the number of tasks that contain hints about the visible text in the answer images, with 11/23 tasks having OCR information. These OCR clues play a critical role in helping participating systems find the correct images, as most of the time teams solved tasks based on them. Across 12 times that Myscéal was one of the three fastest teams, there were eight times that Myscéal found the answers using the OCR feature, which was only implemented for LSC'21. This OCR update indeed came from the feedback of novice users in our experiments (Section 5.1) when they commented that it would be easy to solve Task 3 in Table 3 if they could use OCR to search for "*corona extra*". In addition, we also had some modifications in Myscéal after LSC'20 and to prepare for LSC'21 based on our observations in the user experiments. For example, the map area in our interface was then enlarged to effectively grasp the attention of users since most the novice users did not utilise this helpful feature as they did not realise there was a map in the interface.

Regarding ImageCLEFlifelog'20, since Myscéal was not created to match the evaluation metrics of this challenge, the system could only achieve third place in this competition. However, we considered ImageCLEFlifelog'20 to be a good opportunity to conduct a user study, including the lifelogger as a user for Myscéal. Table 5 showed that the expert user could have a similar score to the lifelogger with significant knowledge about the dataset. However, although knowing the dataset can have an impact on the lifelog retrieval result, this merit is not enough to gain a high score in ImageCLEFlifelog'20 when this competition required searchers to find all relevant moments. This is because these events sometimes cannot be remembered by the data owner due to the massive size of the dataset with nearly 200.000 images. The target of a lifelog retrieval system is the lifelogger since the system is just a tool supporting them to recall a specific event. By having the same scores for both the expert and the lifelogger, it shows the benefits of using Myscéal to retrieve lifelog images since the expert does not know the dataset as well as the lifelogger but understands how the system works. Furthermore, we believe that if the lifelogger, who is already familiar with their own dataset, has enough time to learn how to use Myscéal, they can even achieve a better score.

Through three lifelog retrieval competitions, Myscéal has learned many lessons from which we propose some upgrades for the newest version of Myscéal, called E-Myscéal, to participate in LSC'22 [30]. The updates can be summarised as follows:

- Continuously implementing the OCR feature.
- Updating the user interface that emphasises useful features such as the similar images area.
- Adding the embedding technique to encode both visual images and semantic queries to the same latent space where we can measure the similarity between these multimodal data.

7 Conclusion

We have described Myscéal, which is the current state-of-the-art lifelog retrieval system. Some user experiments have been discussed to offer insights into the system's performance. Myscéal applied the conventional and standard approach in this research field, which is to annotate images using its visual concepts but introducing our own new feature called aTFIDF. This novel feature is introduced with the belief that larger visual objects in an image will be more important than smaller objects. In addition to the search engine, Myscéal comes with a clean and simple user interface to support novice users unfamiliar with this area. We have also shown how Myscéal updated both the back-end engine and the front-end interface through the competitions in which the system participated. Among the 3 lifelog challenges, our Myscéal achieved the first place in LSC'20 and LSC'21 while obtaining a considerable third place in ImageCLEFlifelog'20. This result shows the competitive performance of Myscéal compared to other lifelog retrieval systems.

Declarations

Funding: This publication has emanated from research supported in party by research grants from Science Foundation Ireland Centre for Research Training under grant numbers 18/CRT/6223 and 18/CRT/6224.
Conflicts of Interest: The authors have no competing interests to declare that are relevant to the content of this article.

References

[1] Gurrin, C., Smeaton, A.F., Doherty, A.R., *et al.*: Lifelogging: Personal

big data. Foundations and Trends® in information retrieval **8**(1), 1–125 (2014)

[2] Gurrin, C., Schoeffmann, K., Joho, H., Leibetseder, A., Zhou, L., Duane, A., Nguyen, D., Tien, D., Riegler, M., Piras, L., *et al.*: Comparing approaches to interactive lifelog search at the lifelog search challenge (LSC2018). ITE Transactions on Media Technology and Applications **7**(2), 46–59 (2019)

[3] Dang Nguyen, D.T., Piras, L., Riegler, M., Boato, G., Zhou, L., Gurrin, C.: Overview of imageclef lifelog 2017: lifelog retrieval and summarization (2017)

[4] Gurrin, C., Joho, H., Hopfgartner, F., Zhou, L., Albatal, R.: Overview of ntcir-12 lifelog task. (2016). NTCIR

[5] Tran, L.-D., Nguyen, M.-D., Binh, N.T., Lee, H., Gurrin, C.: Myscéal: An experimental interactive lifelog retrieval system for LSC'20. In: Proceedings of the Third Annual Workshop on Lifelog Search Challenge, pp. 23–28 (2020)

[6] de Oliveira Barra, G., Cartas Ayala, A., Bolaños, M., Dimiccoli, M., Giró Nieto, X., Radeva, P.: 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 (2016)

[7] Le, N.-K., Nguyen, D.-H., Nguyen, V.-T., Tran, M.-T.: Lifelog moment retrieval with advanced semantic extraction and flexible moment visualization for exploration. In: CLEF (Working Notes) (2019)

[8] Duane, A., Gurrin, C., Huerst, W.: Virtual reality lifelog explorer: lifelog search challenge at acm icmr 2018. In: Proceedings of the 2018 ACM Workshop on The Lifelog Search Challenge. LSC'18, pp. 20–23 (2018)

[9] Rossetto, L., Gasser, R., Heller, S., Amiri Parian, M., Schuldt, H.: Retrieval of structured and unstructured data with vitrivr. In: Proceedings of the ACM Workshop on Lifelog Search Challenge. LSC'19, pp. 27–31 (2019)

[10] Leibetseder, A., Schoeffmann, K.: Lifexplore at the lifelog search challenge 2020. In: Proceedings of the Third Annual Workshop on Lifelog Search Challenge. LSC '20, pp. 37–42. Association for Computing Machinery, New York, NY, USA (2020)

[11] Kovalčík, G., Škrhak, V., Souček, T., Lokoč, J.: Viret tool with advanced visual browsing and feedback. In: Proceedings of the Third Annual Workshop on Lifelog Search Challenge. LSC '20, pp. 63–66. Association for

Computing Machinery, New York, NY, USA (2020)

[12] Trang-Trung, H.-P., Le, H.-A., Tran, M.-T.: Lifelog moment retrieval with self-attention based joint embedding model. In: CLEF (Working Notes) (2020)

[13] Alateeq, A., Roantree, M., Gurrin, C.: Voxento 2.0: A prototype voice-controlled interactive search engine for lifelogs. In: Proceedings of the 4th Annual on Lifelog Search Challenge. LSC '21, pp. 65–70. Association for Computing Machinery, New York, NY, USA (2021)

[14] Alam, N., Graham, Y., Gurrin, C.: Memento: A prototype lifelog search engine for LSC'21. In: Proceedings of the 4th Annual on Lifelog Search Challenge. LSC '21, pp. 53–58. Association for Computing Machinery, New York, NY, USA (2021)

[15] Rossetto, L., Baumgartner, M., Gasser, R., Heitz, L., Wang, R., Bernstein, A.: Exploring graph-querying approaches in lifegraph. In: Proceedings of the 4th Annual on Lifelog Search Challenge. LSC '21, pp. 7–10. Association for Computing Machinery, New York, NY, USA (2021)

[16] Nguyen, M.-D., Nguyen, B.T., Gurrin, C.: Graph-based indexing and retrieval of lifelog data. In: International Conference on Multimedia Modeling. MMM'21, pp. 256–267 (2021). Springer

[17] Mejzlík, F., Veselý, P., Kratochvíl, M., Souček, T., Lokoč, J.: Somhunter for lifelog search. In: Proceedings of the Third Annual Workshop on Lifelog Search Challenge. LSC '20, pp. 73–75. Association for Computing Machinery, New York, NY, USA (2020)

[18] Lokoč, J., Mejzlík, F., Veselỳ, P., Souček, T.: Enhanced somhunter for known-item search in lifelog data. In: Proceedings of the 4th Annual on Lifelog Search Challenge. LSC'21, pp. 71–73 (2021)

[19] Tran, L.-D., Nguyen, M.-D., Thanh Binh, N., Lee, H., Gurrin, C.: Myscéal 2.0: A revised experimental interactive lifelog retrieval system for LSC'21. In: Proceedings of the 4th Annual on Lifelog Search Challenge. LSC'21, pp. 11–16 (2021)

[20] Chen, L.-C., Zhu, Y., Papandreou, G., Schroff, F., Adam, H.: Encoder-decoder with atrous separable convolution for semantic image segmentation. In: Proceedings of the European Conference on Computer Vision (ECCV), pp. 801–818 (2018)

[21] Anderson, P., He, X., Buehler, C., Teney, D., Johnson, M., Gould, S., Zhang, L.: Bottom-up and top-down attention for image captioning and visual question answering. In: Proceedings of the IEEE Conference on

Computer Vision and Pattern Recognition, pp. 6077–6086 (2018)

[22] Krishna, R., Zhu, Y., Groth, O., Johnson, J., Hata, K., Kravitz, J., Chen, S., Kalantidis, Y., Li, L.-J., Shamma, D.A., *et al.*: Visual genome: Connecting language and vision using crowdsourced dense image annotations. International journal of computer vision **123**(1), 32–73 (2017)

[23] Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556 (2014)

[24] Mikolov, T., Sutskever, I., Chen, K., Corrado, G.S., Dean, J.: Distributed representations of words and phrases and their compositionality. In: Advances in Neural Information Processing Systems, pp. 3111–3119 (2013)

[25] Lowe, D.G.: Object recognition from local scale-invariant features. In: Proceedings of the Seventh IEEE International Conference on Computer Vision, vol. 2, pp. 1150–11572 (1999)

[26] Lowe, D.G.: Distinctive image features from scale-invariant keypoints. International journal of computer vision **60**(2), 91–110 (2004)

[27] Oram, P.: Wordnet: An electronic lexical database. christiane fellbaum (ed.). cambridge, ma: Mit press, 1998. pp. 423. Applied Psycholinguistics **22**(1), 131–134 (2001)

[28] Luo, H., Wei, H., Lai, L.L.: Creating efficient visual codebook ensembles for object categorization. IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans **41**(2), 238–253 (2011)

[29] Tran, L.-D., Nguyen, M.-D., Nguyen, B.T., Gurrin, C.: An experiment in interactive retrieval for the lifelog moment retrieval task at ImageCLEFlifelog2020. In: CLEF (Working Notes), p. 12 (2020)

[30] Tran, L.-D., Nguyen, M.-D., Thanh Binh, N., Lee, H., Gurrin, C.: E-myscéal: Embedding-based interactive lifelog retrieval system for LSC'22. In: Proceedings of the 5th Annual Lifelog Search Challenge (2022)

[31] Nguyen, T.-N., Le, T.-K., Ninh, V.-T., Tran, M.-T., Thanh Binh, N., Healy, G., Caputo, A., Gurrin, C.: Lifeseeker 3.0: An interactive lifelog search engine for LSC'21. In: Proceedings of the 4th Annual on Lifelog Search Challenge, pp. 41–46 (2021)