DCU and HCMUS at NTCIR-16 Lifelog-4

Thao-Nhu Nguyen
Tu-Khiem Le
Van-Tu Ninh
Dublin City University
Ireland

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

Minh-Triet Tran
Thanh-Binh Nguyen
University of Science, VNU-HCM
Vietnam National University
Ho Chi Minh City
Vietnam

Annalina Caputo
Sinéad Smyth
Graham Healy
Cathal Gurrin
Dublin City University
Ireland

ABSTRACT

In this paper, we present our DCU and HCMUS team's participation in the NTCIR16 Lifelog-4 task by using two different retrieval systems, namely LifeSeeker and Myscéal that were originally introduced in the Lifelog Search Challenge (LSC) and adapted for addressing the Lifelog Semantic Access Task (LSAT). To tackle the task in an automatic manner, both LifeSeeker and Myscéal employed pre-processing techniques as part of the retrieval process, while LifeSeeker further utilised a post-processing step to refine the retrieval results. Regarding the interactive manner, we evaluated Myscéal system by conducting a user study on both expert and novice users in both ad-hoc and known-item-search settings.

KEYWORDS

lifelog, information retrieval, quantified self, personal data

TEAM NAME

DCU and HCMUS

SUBTASKS

Lifelog Semantic Access Task (LSAT)

1 INTRODUCTION

Lifelog retrieval is an application information retrieval but focuses on lifelog data, which are archives of personal experimental multimodal data captured using a range of sensors in daily life. Many lifelog retrieval challenges have been organised to encourage the research community to propose their solutions for this research field [5, 7, 9, 17]. Similar to general information retrieval, lifelog retrieval requires participants to find answers to information needs, which in the case of the current lifelog challenges, is to find one (or more) images relevant to the aforementioned information need.

NTCIR-16 hosted the fourth installation of the NTCIR Lifelog task [25] after successfully organising this task from NTCIR-12 to NTCIR-14 [6–8]. Four teams have introduced their interactive retrieval systems to the LSAT in the previous installation at NTCIR-14. While Ninh et al. [19] from DCU provided a baseline model which utilised faceted filtering and novel navigation through ranked lists, the HCMUS team [11] enriched their system knowledge by obtaining habit-based syntaxes and scene-related syntaxes. Additionally, QUIK (Japan) from Kyushu University [22] has considered the relation between lifelog images and images collected from an online search platform (Google image search) to provide aid to retrieval. To do so, they compute the similarity between lifelog images and images obtained from online sources. Meanwhile, the NTU team [3] implemented pre-trained word embedding models and adopted a probabilistic relevance-based ranking function for retrieval to bridge the semantic gap between textual query and visual concepts.

Following the structure of its preceding iterations and also a standard lifelog retrieval challenge, the NTCIR-16 Lifelog-4 task this year asks participating teams to find all (or as many as possible) lifelog images that are relevant to queries. The task can be solved automatically or interactively. Nevertheless, NTCIR-16 Lifelog-4 task is not similar to the Lifelog Search Challenge (LSC) [9] which is another lifelog retrieval challenge, although both competitions allow interactive systems to solve queries. While the latter considers the search time as a fundamental aspect of the scoring mechanism, the former focuses more on the accuracy of the search engine of systems rather than its retrieval time. In addition, LSC only needs participants to retrieve a single correct image, whilst the NTCIR-16 Lifelog-4 task reflects a more conventional retrieval task, in that it requires seeking all relevant images. Moreover, the Lifelog Semantic Access Task (LSAT) in NTCIR-16 includes 48 queries, which is the largest query bank in existing lifelog-related contests. It consists of 2 subtasks with 24 queries each: (1) ad hoc subtask forms the situation where users try to seek as many targets as possible, and (2) known-item subtask indicates the scenario of locating one or few images matching the given description.

In this work, we participate in the NTCIR-16 Lifelog-4 task and resolve queries using two distinct state-of-the-art lifelog retrieval systems, achieving high performance in the most recent LSC'21 challenge. We modify these systems for the NTCIR16-Lifelog 4 LSAT subtask and examine how well these engines perform in the automatic and interactive tasks, including an interactive experiment with novice users.

2 EXPERIMENT 1 - LIFESEEKER

We begin with a description of the LifeSeeker retrieval system, which had previously been used as an interactive lifelog retrieval system, but which operated in an automatic manner for this challenge.

2.1 LifeSeeker System Overview

LifeSeeker [12], first introduced in the Lifelog Search Challenge in 2019 (LSC'19), is a lifelog interactive retrieval system based on the visual and textual information. Over 3 years of participation in the competition [12, 13, 18], the system underwent multiple rounds of development, adjustment, and enhancement. New functionalities were introduced every year to maintain the compatibility with new benchmarking datasets and also to improve the user experience and search accuracy. Figure 1 illustrates the general architecture of the most recent implementation of LifeSeeker. The visual concepts generated from the Microsoft Computer Vision API [1] are provided by the organisers. Alongside with those features, we expand the syntax collection with more concepts by implementing both the Bottom-Up Attention model [1] and Mask-RCNN [10, 24] pre-trained on the COCO dataset [14]. Regarding metadata enhancement, we not only refine 32 separate categories of semantic locations, but also get the address and annotate labels for city, country based on the given geographic coordinates. Time alignment is necessary to ensure the consistency across lifelogger's wearable sensors. Furthermore, the visible text in the lifelogging images was extracted via Google Vision API [2]. Finally, all the aforementioned annotations are inserted and indexed into both the ElasticSearch [4] and Weighted Bag-of-words models for later retrieval purposes.

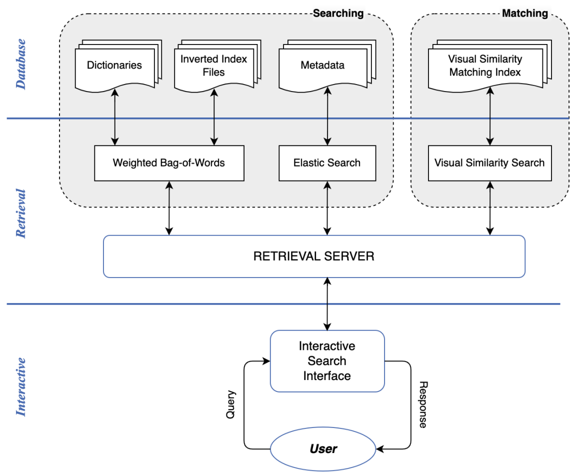


Figure 1: LifeSeeker's system architecture [18]

Inspired by the Google-style text box, the User Interface (UI), shown in Figure 2, is designed for for use by novice users. While the main window consists of the free-text-query box on the top and the result images displayed right below, the expandable box enables users to have more details about time and date of the chosen frame. Besides, LifeSeeker's UI has been developed by offering new navigation options for temporal search and visual similarity exhibition.

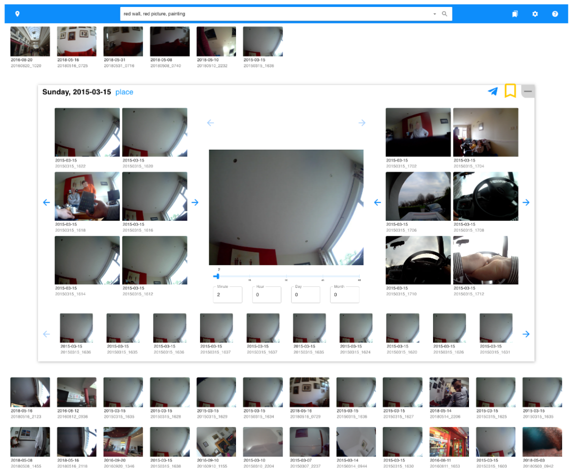


Figure 2: LifeSeeker's User Interface [18]

2.2 Automatic Approach

At the very beginning, LifeSeeker is designed as an interactive retrieval system which requires user communication. For our participation in the automatic task of NTCIR-Lifelog, we integrate new mechanisms for auto-processing query and post-processing results.

Query Processing. From the given full description, we manually generate a simple version consisting of all key information (without any duplication or stopwords). Then, those queries passed to the text parser in order to separate into keywords belonging to 3 different fields: location, time, and concepts. Particularly, while specific place and proper nouns are extracted by matching with a syntax dictionary, we leverage SUtime [2] to get temporal information. Otherwise, those keywords are considered as concepts along with objects, humans. For instance, the query "*I was drinking coffee while waiting in a car repair / sales store in May 2018*" will be parsed into "*repair/sales store*", "*May 2018*" and "*drinking coffee*" corresponding to location, time and concepts field, respectively.

Post-processing. Having no ability to interact with the system, compared to the interactive task, we tried to maximise correctness matching by proposing three different post-processing approaches to return the final ranked lists: **(A1)** Get the top ranked 100 images (where 100 are available), **(A2)** Combine the first 10 images with the 8 temporally closest neighbouring images (4 forwards and 4 backwards) and **(A3)** Combine the first 20 images with the 4 temporally closest images (2 forwards and 2 backwards). This resulted in three runs. In order to examine variance in query generation, we got two different users to generate the single query string for each topic and compare their results, meaning that we submitted six official runs.

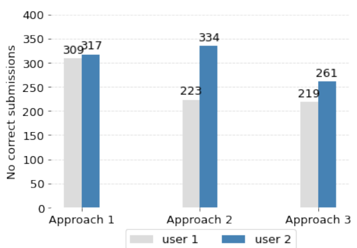
[1] https://azure.microsoft.com/en-us/services/cognitive-services/computer-vision/
[2] https://cloud.google.com/vision/docs/ocr

Table 1: Results of LifeSeeker runs.

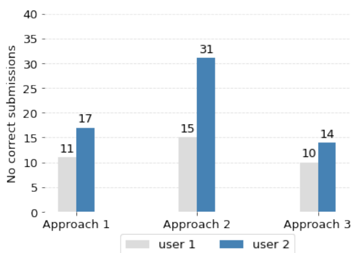
Run	# queries attempted	# queries solved/submitted	# images correct/submitted	MAP	P@5	P@10
U1-A1	48	29/48	320/4629	**0.0299**	**0.0833**	**0.0750**
U2-A1	48	**31**/48	334/4677	0.0211	0.0583	0.0583
U1-A2	48	16/48	238/4530	0.0236	0.0500	0.0583
U2-A2	48	22/48	**365**/4708	0.0237	0.0792	0.0729
U1-A3	48	19/48	229/3714	0.0286	0.0792	0.0667
U2-A3	48	22/48	275/3846	0.0168	0.0583	0.0625

2.3 Result

For the non-interactive task, each of the two expert users (denoted as "**U1**" and "**U2**") constructed one file consisting of 48 information needs, before submitting them to the LifeSeeker search engine. By doing so, we generate 6 submissions whose results are shown in Table 3. Table 1 indicates the result summary of all runs, which highlights the number of queries solved, the number of images correct, the mean average precision (MAP), top-5 precision, and top-10 precision. In general, users using Approach 1 performed the best among all approaches with more than 29 out of 48 queries (60%) having been solved. By using this method, User 1 achieved the best run of LifeSeekerwith the highest MAP score, P@5 and P@10 of 0.0299, 0.0833 and 0.0750, respectively. This means that the implemented temporal-inclusion approach did not bring any positive benefit to the retrieval process, when averaged over all queries. However, when examining the results on a per-user and per approach basis, it becomes apparent that there are some benefits evident (in terms of recall) when implementing the temporal-inclusion process.



(a) Ad hoc subtask



(b) Known-item subtask

Figure 3: Number of correct images from all runs for the ad hoc and known-item subtask

Since the objectives of the 2 subtasks in LSAT are quite different, we further investigate the system's performance of each subtask separately, as shown in Figure 3. The ad hoc subtask has more groundtruth than the known-item one, which explains why users found more positive images for the former compared to those number for the latter (334 and 31 in **U2-A2** case). For both tasks, the expandable post-processing method (Approach 2) found more relevant items for User 2 when compared to Approach 1. This results happen in case neighboring images of correct moments are included in the submissions, which results in a higher number of relevant items in temporal order compared to the conventional ranked list approach. However, the same situation is not correct for User 1 in which using Approach 1 is their best choice for the ad hoc task. The last approach (Approach 3) seems to perform poorly for both 2 tasks of the 2 users.

Overall, it is apparent that the ability of the user to construct an information need from the provided topics, with user 2 constantly outperforming in terms of the number of queries solved. However, it appears that (in most cases) relying on a conventional ranked list is the best approach to take, rather than exploring the temporal inclusion of neighbouring images.

3 EXPERIMENT 2 - MYSCÉAL

The second system we implemented for this task was the Myscéal system, another interactive lifelog retrieval tool that was designed for the LSC challenge. The Myscéal system was the top performing system at both the LSC'20 and LSC'21 challenges.

3.1 System Overview

Figure 4 illustrates the architecture of the MySceal interactive system. Similarly to LifeSeeker, Myscéal employs a concept-based search approach. Visual descriptors, extracted from images, and non-visual metadata such as GPS coordinates, semantic locations, time, and date, are aligned and indexed in the ElasticSearch[4] engine.

To address the shortcomings of using a fixed list of "keyword" concepts (obtained from pre-trained object detectors), Myscéal applies a query expansion process to search queries. This process makes use of ad-hoc regular expression patterns to break down textual queries into different cues, for example, visual cues, location hints, and temporal constraints. Visual cues are expanded word by word using Word2vec[16] and WordNet[20], keeping only the words that appear in the set of fixed, indexed concepts.

Moreover, Myscéal also supports multiple query searches based on their temporal relationship, a key feature of lifelogs. However, at this point, the system can handle up to only three queries: one `before` query, one `main` query, and one `after` query. Here is an example use case for this functionality: *"Find the moment when I am having food before a flight. I have just arrived at the airport after working in the office."*. In this case, the `before` query is "working in the office", the `main` query is "having food at the airport", and the `after` query is "being on a flight".

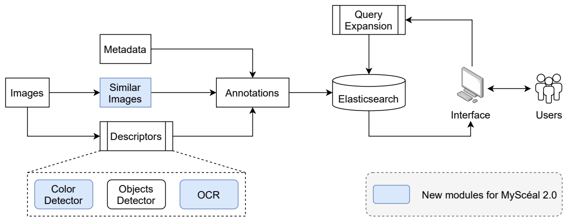


Figure 4: Pipeline of Myscéal.[23]

Since the system was designed to be interactive, several modules were added to assist the user in finding the target images, namely, similar image search, map search, and temporal browsing. These modules are presented in the user interface as seen in Figure 5. Similar scores between images are precomputed using a combination of VGG16[21] and SIFT[15] features. The map section, which occupies a large portion of the interface, allows the user to locate the target location and filter the results using a rectangular boundary of GPS coordinates. It is also useful for showing the location information when the user inspects each image. Temporal browsing is used to examine each lifelog image in its temporal context within the day.

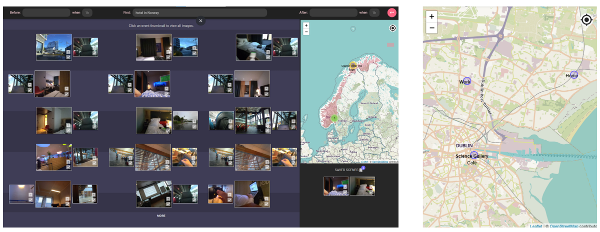


Figure 5: Myscéal user interface.[23]

Myscéal submitted both automatic and interactive runs, which we will now describe.

3.2 Myscéal Automatic Approach

To get the results for each query in the automatic run, we employed an automatic procedure to convert the queries into a suitable format with the least human intervention possible.

- **Temporal splits**: the query cues are split into `before`, `main`, and `after` if applicable;
- **Trimming and changing keywords**: using the result of query expansion, words that do not appear in the indexed keywords are removed or changed into the most sensible word in the expanded list. For example, when entering "My Panda bear, a soft toy who normally resides in my bedroom at home." into the search bar, the query suggestion feedback will highlight words such as "my", "panda", "soft", and "normally", "resides"; and suggest using "bear" for "panda". The final query will be "bear toy in bedroom at home". This is done without looking at any of the search results.

3.3 Interactive Runs

We performed two runs in an interactive manner, one for an expert user (the system designer) and one for novice users. For each query, the user has a bit of time to read the query with all the relevant information before entering the first search query, when the countdown clock starts from 300 seconds to 0. As soon as the user finds a relevant image on the result page, they can submit using the accompanying submit button. If the user submits a thumbnail of a scene (consisting of consecutive similar images) in the temporal browsing view, every image in that scene is submitted at the same time.

Due to our resource limitation, we could only complete the ad-hoc queries, which are the first 24 out of 48 for the novice run. Two novice users were selected, one of whom is the author's acquaintance with no computer science or lifelog background, and the other is a Ph.D. candidate in computer science. Both were offered sample queries to become familiar with the system navigation. The first user finished the first 10 queries and the second the rest. All results are concatenated into one single run before being submitted for judgement.

For the expert run, one of the system designers used the Myscéal system to submit the runs, which is labeled as *expert** and the expert processed all 48 topics.

To provide a comparable result for the novice run, we also submitted shortened versions of the automatic run taking into account only the ad-hoc queries. We address this run by *auto**. Also we compare only the ad-hoc queries from the expert run also.

3.4 Results

The results of three different runs of Myscéal are detailed in Table 2. We used the interactive expert run as a reference for comparison as Myscéal achieved the state-of-the-art performance in LSC'21 in this manner.

Out of 48 queries, the expert run submitted images in 44 queries, 41 of which were considered solved; in other words, the results of 41 queries contained at least one ground-truth image. Meanwhile, the automatic run could always find something for every query, but only solve 28 queries. Of the 4,546 images submitted, 336 were positive. Despite having a similar number of correct images to the LifeSeeker's runs, the automatic run of Myscéal achieved higher scores in all metrics in Table 4. This is likely caused by Myscéal's approach of using temporal events and the keyword changes suggested by the query expansion process. Furthermore, because of the different characteristics of the ad-hoc and known-item subtasks, we compare the full automatic run and the shortened one in Table 2 and Table 4. Due to the fact that ad-hoc queries have more ground-truth images, most of the correct submissions come from these queries (337 out of 366 images), and all precision scores of this run are moderately higher. However, this caused a significant decrease in the MAP score, from 0.1366 to 0.0674.

Table 2: Results of Myscéal's runs.

Run	Type	#queries attempted	#queries solved/submitted	#images correct/submitted	MAP	P@5	P@10
Expert	Interactive	48	41/44	783/1646	0.3980	0.4818	0.4614
Auto	Automatic	48	28/48	366/4546	0.1366	0.1917	0.1521
Expert*	Interactive	24	22/23	681/1484	0.2234	0.5478	0.5826
Auto*	Automatic	24	16/24	337/2242	0.0674	0.2916	0.2291
Novice	Interactive	24	19/22	430/1082	0.1302	0.4909	0.4273

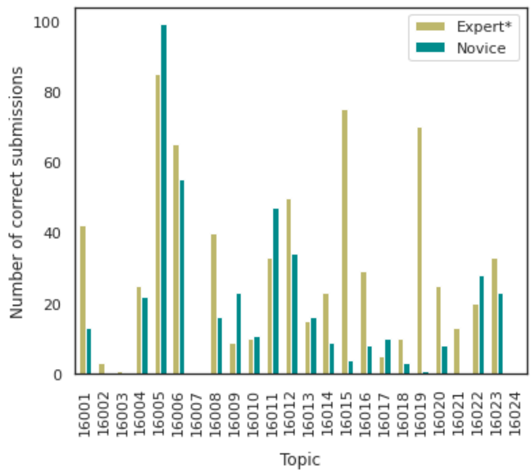


Figure 6: Number of correct submissions for each topic from the expert* and novice runs.

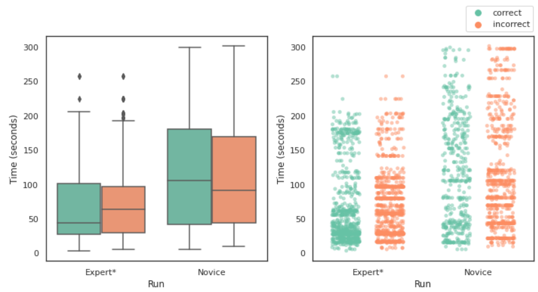


Figure 7: Time elapsed of each image submission from the expert* and novice runs.

To assess the novice run, we examine the shortened (24 topic) *expert** run. The performance of the novice run is lower than that of the *expert**, which is expected. The number of queries submitted is one query short of that of the *expert** run. However, only 19 queries are considered to be solved. The novice users submitted far fewer images than the expert user, with similar accuracy rates: approximately 45% and 40% for the *expert** and novice runs, respectively. As Figure 6 shows, the performances of both runs are competitive. The expert managed to find more correct images in 15 queries, fewer in 7 queries, and the same amount (equal none) in 2 queries (16007 and 16024) compared to the novice run.

Since trec-eval does not take into account submission time, we further analyse the novice run by illustrating the time elapsed of each image submission in Figure 7. Generally, there are no significant differences in speed when comparing correct and incorrect submissions. The expert user tends to submit closer to the beginning, whereas the novice submission times are spread out along the allotted time. This could be explained by the level of familiarity with the system.

Furthermore, when comparing the novice run with the *auto** run, the increase in the scores, especially those related to Precision, could be due to human judgement (which is assumed to be more accurate). The novice users could solve 3 more queries. Nevertheless, the difference in the number of correct images, covering 24 queries, is only 93 images.

4 CONCLUSION

For the NTCIR-16 Lifelog-4 this year, we conducted 2 experimental studies using two separate retrieval systems, LifeSeeker and Myscéal, which aims to evaluate the systems' efficiency for both automatic and interactive manner. By implementing some query-related processing techniques, we resolve more than 58% of the queries on the automatic task. In terms of the interactive subtask, Myscéal showed that the novice user underperformed the experts as measured by accuracy rate (5% lower). Since both systems were initially designed as interactive systems and adapted to partake as automatic systems, it is likely that an optimised automatic system will outperform these results.

5 APPENDIX

ACKNOWLEDGEMENT

This publication has emanated from research supported in party by research grants from Irish Research Council (IRC) under Grant Number GOIPG/2016/741, Science Foundation Ireland under grant numbers 13/RC/2106, 13/RC/2106_P2, SFI/12/RC/2289_P2, 18/CRT/6223, 18/CRT/6224 and SFI/13/RC/2106_P2 and co-funded by the European Regional Development Fund.

REFERENCES

[1] Peter Anderson, Xiaodong He, Chris Buehler, Damien Teney, Mark Johnson, Stephen Gould, and Lei Zhang. 2018. Bottom-Up and Top-Down Attention for Image Captioning and Visual Question Answering. In *CVPR*.
[2] Angel X. Chang and Christopher D. Manning. 2012. SUTime: A library for recognizing and normalizing time expressions. In *LREC*.

Table 3: All results of all runs submitted by LifeSeeker. We omit precision scores P@K where $k > 100$ as the limit of submissions for each query is 100. "Ux - Ay" stands for "User x using approach y".

Run	LifeSeeker (U1-A1)	LifeSeeker (U2-A1)	LifeSeeker (U1-A2)	LifeSeeker (U2-A2)	LifeSeeker (U1-A3)	LifeSeeker (U2-A3)
num_q	48	48	48	48	48	48
num_ret	4629	4677	4530	4708	3714	3846
num_rel	2993	2993	2993	2993	2993	2993
num_rel_ret	320	334	238	365	229	275
map	0.0299	0.0211	0.0236	0.0237	0.0286	0.0168
gm_map	0.0007	0.0007	0.0001	0.0003	0.0002	0.0003
Rprec	0.0351	0.0380	0.0346	0.0410	0.0253	0.0306
bpref	0.1634	0.2014	0.1184	0.1493	0.0992	0.1373
recip_rank	0.1338	0.0971	0.1468	0.1475	0.1207	0.1219
iprec_at_recall_0.00	0.1421	0.1186	0.1491	0.1570	0.1290	0.1379
iprec_at_recall_0.10	0.0621	0.0650	0.0550	0.0737	0.0602	0.0625
iprec_at_recall_0.20	0.0341	0.0296	0.0449	0.0348	0.0389	0.0321
iprec_at_recall_0.30	0.0336	0.0255	0.0427	0.0314	0.0382	0.0238
iprec_at_recall_0.40	0.0335	0.0253	0.0424	0.0247	0.0171	0.0055
iprec_at_recall_0.50	0.0142	0.0083	0.0221	0.0225	0.0170	0.0055
iprec_at_recall_0.60	0.0128	0.0050	0.0018	0.0023	0.0150	0.0055
iprec_at_recall_0.70	0.0128	0.0050	0.0018	0.0023	0.0150	0.0025
iprec_at_recall_0.80	0.0128	0.0050	0.0018	0.0023	0.0150	0.0025
iprec_at_recall_0.90	0.0128	0.0050	0.0018	0.0023	0.0150	0.0025
iprec_at_recall_1.00	0.0128	0.0050	0.0018	0.0023	0.0150	0.0025
P_5	0.0833	0.0583	0.0500	0.0792	0.0792	0.0583
P_10	0.0750	0.0583	0.0583	0.0729	0.0667	0.0625
P_15	0.0750	0.0611	0.0514	0.0806	0.0611	0.0681
P_20	0.0740	0.0656	0.0490	0.0844	0.0677	0.0688
P_30	0.0681	0.0618	0.0535	0.0792	0.0646	0.0708
P_100	0.0667	0.0696	0.0496	0.0760	0.0477	0.0573

[3] Chia-Chun Chang, Min-Huan Fu, Hen-Hsen Huang, and Hsin-Hsi Chen. 2019. An Interactive Approach to Integrating External Textual Knowledge for Multimodal Lifelog Retrieval. 41–44. https://doi.org/10.1145/3326460.3329163

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

[5] 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

[6] Cathal Gurrin, Hideo Joho, Frank Hopfgartner, Liting Zhou, and Rami Albatal. 2016. Overview of NTCIR-12 Lifelog Task. NTCIR.

[7] 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.

[8] Cathal Gurrin, Hideo Joho, Frank Hopfgartner, Liting Zhou, V-T Ninh, T-K Le, Rami Albatal, D-T Dang-Nguyen, and Graham Healy. 2019. Overview of the NTCIR-14 lifelog-3 task. In *Proceedings of the 14th NTCIR conference.* NII, 14–26.

[9] 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.

[10] Kaiming He, Georgia Gkioxari, Piotr Dollár, and Ross B. Girshick. 2017. Mask R-CNN. *CoRR* abs/1703.06870 (2017). arXiv:1703.06870 http://arxiv.org/abs/1703.06870

[11] Nguyen-Khang Le, Minh-Triet Tran, Hien Nguyen, Hoang Trung Hieu, Thanh-An Nguyen, Trng Đt, Tung Dinh, Quc Lng, Viet-Khoa Vo-Ho, and Vinh-Tiep Nguyen. 2019. Smart Lifelog Retrieval System with Habit-based Concepts and Moment Visualization. 1–6. https://doi.org/10.1145/3326460.3329155

[12] 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* (Ottawa ON, Canada) *(LSC '19)*. Association for Computing Machinery, New York, NY, USA, 37–40. https://doi.org/10.1145/3326460.3329162

[13] Tu-Khiem Le, Van-Tu Ninh, Minh-Triet Tran, Thanh-An Nguyen, Hai-Dang Nguyen, Liting Zhou, Graham Healy, and Cathal Gurrin. 2020. *LifeSeeker 2.0: Interactive Lifelog Search Engine at LSC 2020.* Association for Computing Machinery, New York, NY, USA, 57–62. https://doi.org/10.1145/3379172.3391724

[14] 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

[15] 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.

[16] 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.

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

[18] 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* (Taipei, Taiwan) *(LSC '21)*. Association for Computing Machinery, New York, NY, USA, 41–46. https://doi.org/10.1145/3463948.3469065

[19] Van-Tu Ninh, Tu-Khiem Le, Liting Zhou, Graham Healy, Kaushik Venkataraman, Minh-Triet Tran, Duc Tien Dang Nguyen, Sinead Smyth, and Cathal Gurrin. 2019. *A Baseline Interactive Retrieval Engine for Visual Lifelogs at the NTCIR-14 Lifelog-3 Task.* 29–41. https://doi.org/10.1007/978-3-030-36805-0_3

Table 4: All results of all runs submitted Myscéal. We omit the precision scores P@K where $k > 100$ as the limit of submissions for each query is 100. Myscéal Expert* and Myscéal Auto* runs only take into account ad hoc queries, that is, the first 24 queries.

Run	Myscéal Auto	Myscéal Expert	Myscéal Expert*	Myscéal Auto*	Myscéal Novice
num_q	48	44	23	24	22
num_ret	4546	1646	1484	2242	1082
num_rel	2993	2937	2748	2795	2739
num_rel_ret	366	783	681	337	430
map	0.1366	0.3980	0.2234	0.0674	0.1302
gm_map	0.0019	0.1351	0.0949	0.0025	0.0200
Rprec	0.1601	0.4189	0.3144	0.1035	0.1796
bpref	0.2195	0.5281	0.3278	0.1252	0.1796
recip_rank	0.3622	0.6818	0.6472	0.4388	0.5899
iprec_at_recall_0.00	0.7797	0.8550	0.7985	0.4826	0.6949
iprec_at_recall_0.10	0.6645	0.5406	0.5970	0.1776	0.4126
iprec_at_recall_0.20	0.5677	0.3697	0.4222	0.1001	0.3142
iprec_at_recall_0.30	0.5157	0.3169	0.3436	0.0982	0.1934
iprec_at_recall_0.40	0.4600	0.2727	0.2739	0.0810	0.1062
iprec_at_recall_0.50	0.3888	0.2727	0.1428	0.0733	0.0450
iprec_at_recall_0.60	0.3452	0.1591	0.1392	0	0
iprec_at_recall_0.70	0.3054	0.1591	0.1028	0	0
iprec_at_recall_0.80	0.2444	0.1591	0.0678	0	0
iprec_at_recall_0.90	0.2090	0.1591	0	0	0
iprec_at_recall_1.00	0.2090	0.1591	0	0	0
P_5	0.1917	0.4818	0.5478	0.2917	0.4909
P_10	0.1521	0.4614	0.5826	0.2292	0.4273
P_15	0.1431	0.3985	0.5246	0.2250	0.3970
P_20	0.1271	0.3580	0.4826	0.2083	0.3864
P_30	0.1083	0.3098	0.4449	0.1861	0.3409
P_100	0.0762	0.1780	0.2961	0.1404	0.1955

[20] 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.

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

[22] Tokinori Suzuki and Daisuke Ikeda. 2019. Smart Lifelog Retrieval System with Habit-based Concepts and Moment Visualization.

[23] 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.

[24] Yuxin Wu, Alexander Kirillov, Francisco Massa, Wan-Yen Lo, and Ross Girshick. 2019. Detectron2. https://github.com/facebookresearch/detectron2.

[25] Liting Zhou, Cathal Gurrin, Graham Healy, Hideo Joho, Thanh-Binh Nguyen, Rami Albatal, and Frank Hopfgartner. 2022. Overview of the NTCIR-16 Lifelog-4 Task. In *Proceedings of the 16th NTCIR Conference on Evaluation of Information Access Technologies (NTCIR-16)*. Tokyo, Japan.