NTCIR 17 Conference: Proceedings of the 17th NTCIR Conference on Evaluation of Information Access Technologies, December 12-15, 2023, Tokyo, Japan
DOI: https://doi.org/10.20736/0002001284

MemoriEase at the NTCIR-17 Lifelog-5 Task

Quang-Linh Tran
ADAPT Centre, School of Computing,
Dublin City University
Dublin 9, Ireland
linh.tran3@mail.dcu.ie

Binh Nguyen
University of Science
Ho Chi Minh City, Vietnam
ngtbinh@hcmus.edu.vn

Gareth J. F. Jones
ADAPT Centre, School of Computing,
Dublin City University
Dublin 9, Ireland
gareth.jones@dcu.ie

Cathal Gurrin
ADAPT Centre, School of Computing,
Dublin City University
Dublin 9, Ireland
cathal.gurrin@dcu.ie

ABSTRACT

We present the MemoriEase retrieval system used for out participation in the NTCIR Lifelog-5 Task. We report our method to address the lifelog retrieval problem and discuss our official results of the MemoriEase at Lifelog-5 task. We originally introduced the MemoriEase system for the Lifelog Search Challenge (LSC) as an interactive lifelog retrieval system. We have modified it to an automatic retrieval system to address the NTCIR Lifelog-5 Task. We propose the BLIP-2 model as the core embedding model to retrieve lifelog images from textual queries. The open-sourced Elasticsearch search engine serves as the main engine in the MemoriEase system. Some pre-processing and post-processing techniques are applied to adapt this system to an automatic version and improve the accuracy of retrieval results. Finally, we discuss the results of the system on the task, some limitations of the system, and lessons learned from participating in the Lifelog-5 task for further improvements for the system in the future.

KEYWORDS

lifelog, information retrieval, personal data

TEAM NAME

MEMORIEASE

SUBTASKS

Lifelog Semantic Access SubTask - LSAT
Lifelog Insight SubTask - LIT

1 INTRODUCTION

A Lifelog is an archive containing data captured in relation to the daily life of a lifelogger [7]. This data typically contains egocentric images and videos, geospatial information, psychological and physiological metrics, et cetera. There are a wide range of applications for which an individual's lifelog may provide valuable input, including for example, tracking health [3], and aiding memory recall [2] to retrieve moments in the past [15, 18]. Since the volume of data in a lifelog rapidly becomes enormous, e.g over 1500 images per day can typically be proactively captured of the lifeloggers activities, organizing and retrieving useful data from the lifelog is challenging [7]. The Personal Lifelog Organisation & Retrieval (Lifelog-5) [20] is a subtask in NTCIR-17 which provides a competitive benchmark for organizing and retrieving lifelog images. Lifelog-5 is the first time the MemoriEase system has taken part in this task with an enhanced retrieval system from Lifelog Search Retrieval 2023 (LSC'23) [18].

The NII Testbeds and Community for Information Access Research (NTCIR) is a regular research benchmark held on an 18 months cycle, which providers researchers with large-scale evaluations of AI technologies in information retrieval (IR) and Natural Language Processing (NLP). NTCIR-17 encompasses several core tasks and pilot tasks. Lifelog-5 [20] is one of the core tasks after at NTCIR-17 following the successful hosting of this task from NTCIR-12 to NTCIR-16 [4–6, 21]. Lifelog-5 is organized to encourage sustained collaborative research into ad-hoc retrieval and motivate research into the new research topic of question answering from a Lifelog. The introduction of question answering in this edition of the NTCIR lifelog track, along with the latest LSC, seeks to extend the scope of research into the exploitation of lifelogs. Question answering is a challenging topic, but also an exciting task to address within Lifelog-5. In the previous edition of Lifelog at NTCIR-16, several teams participated in the Lifelog-4 track [21]. Team DCU and HCMUS [12] participated with two systems, namely LifeSeeker [13], and Myscéal [16], in both automatic and interactive manners. While both systems used the conventional concept-based search approach, the Myscéal system achieved a better result than LifeSeeker. In addition, Myscéal also unofficially participated interactively in an experiment for expert and novice users. The results from expert users are the highest compared to novice users and automatic run, showing the benefit of employing an expert user. Against the conventional concept-based approach, team DCU with two systems, DCUMemento and DCUVOX [1], used the image-text embedding-based approach by applying various CLIP models [14] to build their search backend. Another team that participated in Lifelog-4 is THUIR [8], which created an interactive Lifelog search engine with a multi-functional and flexible human feedback mechanism for result retrieval. They submitted both unofficial and official runs to Lifelog-4. The unofficial run achieved the highest score. These previous systems show that the embedding-based approach emerges as an excellent approach to performing lifelog retrieval. In this Lifelog task edition, we continue to improve the retrieval system by applying the latest BLIP-2 embedding models [9] to further exploit the embedding-based approach for lifelog retrieval.

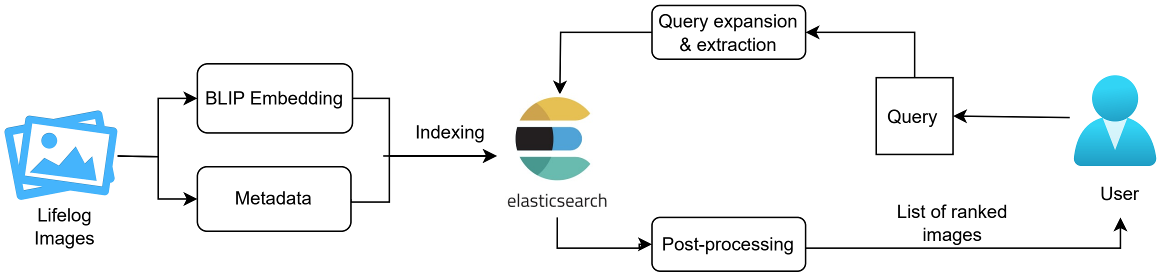

Figure 1: MemoriEase Overview

To take part in the NTCIR-Lifelog-5 this year, we adapted the interactive MemoriEase system from LSC'23 [18] to be suitable for the automatic requirement of Lifelog-5. The most significant improvement is the query pre-processing, in which we apply automatic filtering and query expansion steps to retrieve more relevant results. It can be beneficial in the ad-hoc task, which requires a large number of results. In addition, the quality of the indexed database is enhanced significantly with a new approach for low-quality image removal. An ensemble approach is proposed to remove blurred images and proves high efficiency. We use a post-processing step to optimize the retrieval results and improve accuracy. It is hoped that the MemoriEase system, with many improvements, will show high performance in the NTCIR-Lifelog-5 task.

The structure of the paper is as follows: detailed information about constructing the MemoriEase system is provided in section 2. Section 3 describes query analysis from the committee of NTCIR-17 Lifelog-5, the official results, and some error analysis from the query. Finally, the paper concludes with a summary of the system and suggestions for future improvements in section 4.

2 METHODS

In this section, we provide our method to construct the MemoriEase retrieval system to address the Lifelog-5 task. This system is an enhanced version of the previous MemoriEase system used at LSC'23 with some modifications in data processing and result processing. Figure 1 shows an overview of the system. The dataset providing, including lifelog images, metadata, and visual concepts are processed and indexed into Elasticsearch, as described in subsection 2.1. The process for retrieval image retrieval from the indexed data is presented in subsection 2.2. Subsection 2.3 introduces the new processing techniques enhanced from the previous version of the system.

2.1 Data Processing and Indexing

The dataset provided by the task organizers comprises over 725k images captured over 18 months, ranging from January 2019 to June 2020. This dataset is the same as that used for LSC'23. Along with the lifelog images are the metadata and the visual concepts. The metadata includes the time, physical activities, biometrics, and locations. Meanwhile, the visual concepts are extracted from the non-redacted version of the visual dataset by using the Microsoft Computer Vision API [1]. There are several steps applied to process and index the lifelog images and metadata, including blurred image removal, BLIP-2 visual embedding extraction for images and extraction, cleaning & enhancement for metadata. The full processing and indexing procedure of the dataset is shown in Figure 2.

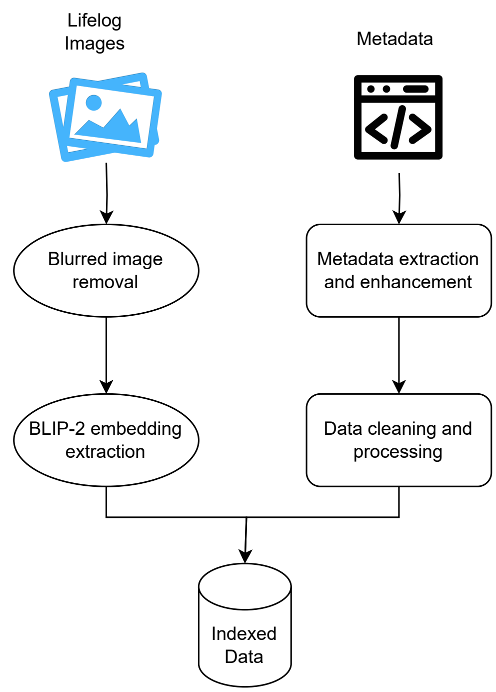

Figure 2: Data processing and indexing process

Because of the nature of lifelog images, which is egocentric with the camera for lifelogging worn on the chest or wrist, there are many blurred and low-quality images in the dataset. By removing these images, the search inference time can be improved thanks to the reduced indexed database size. We apply the method of edge weight summation to discard images with low edge weight. The previous system [18] used this method and proved the efficiency of removing blurred images and preventing them from appearing in the results. This method calculates all the edge weights of objects in the image and sums them up to create a factor to measure the blur degree of images. Blurred images with unclear objects have a smaller sum of edge weights than clear, good-quality images. The

[1]https://azure.microsoft.com/en-us/products/ai-services/ai-vision/

location and semantic name of the metadata also are enhanced through the approach proposed by Tran et al [17]. Compared to the earlier system used for participation in LSC'23, a modification in this system is that we removed the event segmentation in this dataset. The reason is to improve the returned results for ad-hoc tasks. With no event segmentation, the results are more relevant images to the query.

To improve the accuracy of the results, we integrated more filters using the metadata field. Three new filters, namely semantic name, hour, and is_weekend are added, along with our existing filters, including date, city, and weekday. The automatic manner of the Lifelog-5 task does not require search time, adding these filters increases the search time but improves the accuracy. There are some complex queries that require many layers of filters to get accurate results. For example, "Find the moment when I was in the taxi before 6 am, on the way to an airport". This query requires a filter of images before 6am and the semantic name airport to get an accurate result. To deal with this type of complex query, we enhanced the filters to capture more relevant information to filter out non-relevant images.

A BLIP-2 model [9] extracts the embedding from the images. This version is the latest version of BLIP models, compared to the BLIP [10] in the previous MemoriEase system. BLIP-2 is a generic and efficient multi-modal pre-training model that bootstraps vision-language pre-training from off-the-shelf frozen pre-trained image encoders and frozen large language models. It achieved state-of-the-art performance on various tasks related to multi-modality, such as zero-shot Image-to-Text generation, image captioning, visual question-answering, and especially in image-text retrieval, outperformed previous multi-modal pre-trained models such as BLIP [10], ALBEF [11], and CLIP [14]. The original embedding size of the BLIP-2 model is 768 dimensions, but to improve the inference time, we only get the projected embedding with a dimension of 256. An embedding of 256 dimensions for each image is extracted and combined with other metadata to create a dataset for indexing to Elasticsearch[2]. The total number of data points for indexing after all pre-processing steps is over 650K.

2.2 Search Engine

Elasticsearch is the core search engine in the MemoriEase system. It is a powerful, open-source search and analytics engine designed to efficiently store, search, and analyze large volumes of data in various forms. Elasticsearch provides real-time search capabilities and is widely used for a variety of applications, from log and event data analysis to full-text search in documents. With the development of Artificial Intelligence (AI), Elasticsearch now provides vector search features, which inherit the vector from AI models. By converting textual query and lifelog images into vector representations (numerical representations of the meaning of the text and image), the search engine can comprehend the meaning and context of the query and image to perform a better search. Compared to a traditional keyword-based search, instead of relying on keywords of query and visual concepts of images, vectors enable the process of embedding query and images to numerical values. The BLIP-2 model extracts the embedding of the query and feeds it as the input to the Elasticsearch indices, accompanied by other filters, to perform vector search with indexed image embedding and metadata. The database first filters out all irrelevant images by the metadata extracted from the query.

A noteworthy point in the search engine is that it performs quickly even when the dataset is not grouped by event. The search time is less than several seconds when performing the cosine similarity calculation with all the embedding in the Elasticsearch database. This is because Elasticsearch uses an approximate k-nearest neighbor (kNN) algorithm with a limited number of candidates on each shard. It computes the similarity of these candidate vectors to the query vector, selecting the k most similar results from each shard, and then merges the results from each shard to return the global top k nearest neighbors or in other words, the top relevant images to the query. The metadata filters also help to reduce the search time since they narrow the search space of the embedded images to only filtered images.

2.3 New Processing Techniques

Several new processing techniques are included the MemoriEase system used in our NTCIR-15 task participation compared to the previous one in both the pre-processing and post-processing steps. These new processes are shown in figure 3 within blue and green rectangles. Firstly, a query expansion step is used to paraphrase the original query and create new queries. This aims to diversify the meaning of the original query to avoid any missing results due to the limited words in the original query. This process is performed using ChatGPT[3], thanks to its superior understanding of diverse fields. It has already shown its in several previous research [19]. The prompt used to create a new query is 'You are a helpful paraphrasing tool, create a new query from this query but do not add any more information.'. After collecting several paraphrased queries, we perform the search in parallel using both the original query and these new queries. These queries only share the same filters, while embedding vectors are different. Each query creates a list of relevant results. We combine the results using a weighted averaging of the relevance score. The weight for the original query is 0.5, while the weight for paraphrased queries is 0.5, divided equally by each query. This weight is chosen to ensure the importance of the original query. A post-processing step is applied to reorder the results by the calculated score. Secondly, We also enhance the temporal search by automating the temporal extraction process. To be specific, the system can automatically extract the main event, the previous and next event of the query, and as a result, it can help the model to better understand the context and retrieve more accurate results.

3 EXPERIMENTS

This section discusses the official results of MemoriEase at NTCIR-17 Lifelog-5 and some analysis of query and error to find the limitations of the system.

3.1 Query Analysis

There are a total of 41 queries in the NTCIR-17 Lifelog5, comprising 17 Ad-hoc queries and 24 known-item queries. All queries have a title, which is a short title to describe the main content of the query

[2]https://www.elastic.co/

[3]https://chat.openai.com/

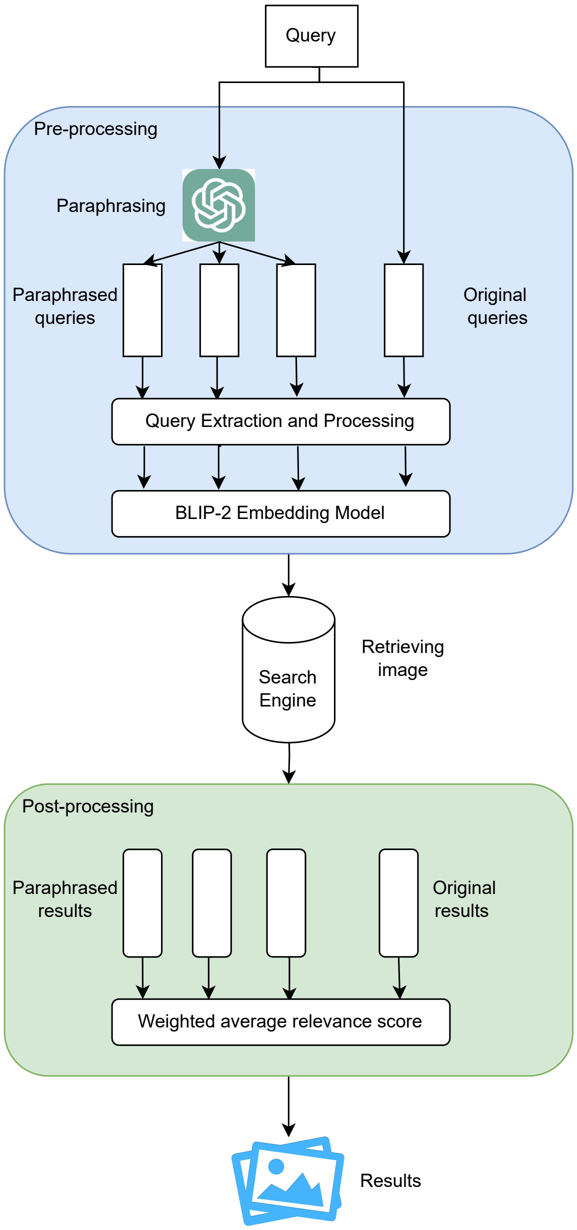

Figure 3: New data pre-processing and post-processing

such as 'Eating Avocado', and 'Taking Medication'. There is also a description for the query to provide more information needed to find the relevant images. For example, with the title 'Eating Avocado', the description is 'Find examples of when I was eating avocado for breakfast.'. The description provides more information about the query by adding who eats the avocado and when the event happens. The description helps to narrow the search space. Another important attribute is narrative, which specifies more about the query to ensure the system finds only relevant images. For example, the narrative for the title 'Eating Avocado' is 'To be relevant, the images must show the lifelogger was eating avocado. Other fruits are not relevant'. This narrative indicates that other fruits are not relevant, which can help to filter out other images containing fruits.

MAP	0.2713
GM_MAP	0.0283
Bpref	0
Rprec	0.29
Recip Rank	0.6197

Table 1: The performance of the MemoriEase in various metrics.

While the ad-hoc queries mainly contain general information, the known-item queries are more specific. For example, the query 'Mother Mary prays for us' with the description 'Find the moment when Mother Mary was praying for us. A religious image in front of a window on a Saturday morning' and the narrative 'A religious image in front of a window on a Saturday morning in Dublin City University after coffee and before a second coffee in a different cafe' provides a large amount of information about the moment to find. To be specific, the query describes the moment with a religious image that is in front of a window. In addition, the query also indicates that it happened on Saturday morning at Dublin City University. Moreover, there is also temporal information in this query, which is drinking coffee before and after seeing the religious picture in different cafes. From the provided information, we can manually extract the necessary information, and perform a search with relevant filters.

After analyzing all the queries in the NTCIR-17 Lifelog5, there are several difficult queries that can challenge the retrieval system to find the correct answers. These queries are shown in Table 3. These queries contain some objects that can be difficult to recognize such as medication, vitamins, or oysters because of the size of objects. In addition, these queries also require information from multiple sources such as images and music metadata, to find the exact moment. Finally, some queries include personal names, which require prior knowledge to find the relevant images.

3.2 Results

We submitted a total of 4100 images for 41 queries, corresponding to 100 images per query. Meanwhile, the total number of relevant images for all queries is 7494 images, which is nearly double the number of results we submitted. There are a total of 651 submitted images considered correct, accounting for approximately 15.88% of the number of submitted images and 8.68% of the relevant images. More metrics about the results are shown in Table 1 and 2.

In Table 1, 5 metrics are shown, including Mean Average Precision (MAP), Geometric Mean Average Precision (GM_MAP), Binary Preference (Bpref), R Precision (Rprec), and Reciprocal Rank (Recip Rank). The MemoriEase system achieved a MAP of 0.2713, which is relatively low because of the small number of relevant retrieved images. Meanwhile, the GM_MAP is only 0.0283 and the Bpref is relatively 0, indicating the low performance of the MemoriEase on resolving the queries. The Rprec is 0.29 and the Reciprocal Rank is 0.6197, suggesting that relevant images are found within the top 15% of the retrieved list.

Cutoff	Precision
@5	0.3707
@10	0.3219
@15	0.2878
@20	0.2621
@30	0.2496
@100	0.1588

Table 2: Precision at different cutoffs

Table 2 delves deeper into the order of retrieved images by measuring the Precision and normalized discounted cumulative gain (NDCG) at different cutoffs, from top 5 to top 100. The precision at 5 (P@5) is 0.3707, indicating that approximately 37.07% of the top 5 retrieved images are relevant. However, when increasing the cutoff, the precision only grows slowly with 0.2878 P@20, 0.2496 P@30, and 0.1587 P@100. This increase shows that most of the relevant images are on the top of the retrieved list. In addition, for ad-hoc queries, the retrieved list is only relevant for top results and the quality of low-ranked images is not high.

3.3 Error Analysis

This subsection provides examples of failed queries resolved by MemoriEase and proposes approaches to improve it. For some challenging queries from Table 3, we use query 17002 - Taking Medication and 17033 - Oyster to do a query analysis. With the ad-hoc query taking medication, we use the description "Find examples of when I was taking medication or vitamins at home" to perform a full-text search. In addition, we add a filter of semantic names with the value 'Home' to only search relevant images captured at home. The results of the search can be seen in Figure 4. As we can see, the diversity of retrieved images is enormous. There are no similar images in the top 15 displayed images, except for the images of the top 1 and top 15, which are considered as correct images. Other images contain other objects which is not medicine or vitamins but look similar such as battery, chewing gum, etc. In addition, there are also some images of medicine packaging boxes. These images are not correct because there is no evidence of taking medicine. To overcome this issue, we can integrate relevant feedback to refine the query to find more relevant images by choosing the correct images and research.

Query 17033 - Oyster is challenging due to the relatively small objects of oysters in the image. To perform the search for this known-item query, we use the main event search with the full-text query 'Find the moment the lifelogger was eating ouster at a BBQ party with very few people' and a temporal search for the previous event is "Drive to home". The result of the search for this query is depicted in figure 5. There is no image of eating oysters, only one image of grilling oysters at 19:06 2019-06-22. The previous event 'Driving to Home' helps the search engine narrow down the search space significantly, in which nearly all of the results contain the driving images before the BBQ party, except for one result. There are also other BBQ images but there is prawn rather than oysters. Moreover, there are 2 images of parties outside with a lot of food but no oysters are seen in the image. To perform a better search for this query, more information such as time or location should be specified to have a better result.

4 CONCLUSIONS

In conclusion, this paper presents the MemoriEase retrieval system, transitioning from an interactive lifelog retrieval tool to an automated system tailored for the NTCIR Lifelog-5 Task. By leveraging the BLIP-2 model as its central embedding model, the MemoriEase system demonstrated efficiency in lifelog image retrieval through textual queries. Incorporating new pre-processing and post-processing techniques further enhances the system's precision of retrieval outcomes. The performance of the MemoriEase system on the NTCIR-17 Lifelog-5 this year is relatively low, especially in ad-hoc queries. For future development, we will continue to enhance the automatic system by upgrading the embedding model and integrating the relevance feedback technique to improve the relevance of returned results.

ACKNOWLEDGMENTS

This research was conducted with the financial support of Science Foundation Ireland at ADAPT, the SFI Research Centre for AI-Driven Digital Content Technology at Dublin City University [13/RC/2106_P2]. For the purpose of Open Access, the author has applied a CC BY public copyright license to any Author Accepted Manuscript version arising from this submission.

REFERENCES

[1] Naushad Alam, Ahmed Alateeq, Yvette Graham, Mark Roantree, and Cathal Gurrin. 2022. DCU at the NTCIR16 Lifelog-4 Task. In *Proceedings of The 16th NTCIR Conference on Evaluation of Information Access Technologies (Tokyo, Japan). National Institute of Informatics.*

[2] Mariona Carós, Maite Garolera, Petia Radeva, and Xavier Giro-i Nieto. 2020. Automatic Reminiscence Therapy for Dementia. In *Proceedings of the 2020 International Conference on Multimedia Retrieval* (Dublin, Ireland) *(ICMR '20)*. Association for Computing Machinery, New York, NY, USA, 383–387. https://doi.org/10.1145/3372278.3391927

[3] Chelsea Dobbins, Reza Rawassizadeh, and Elaheh Momeni. 2017. Detecting physical activity within lifelogs towards preventing obesity and aiding ambient assisted living. *Neurocomputing* 230 (2017), 110–132.

[4] Cathal Gurrin, Hideo Joho, Frank Hopfgartner, Liting Zhou, and Rami Albatal. 2016. Overview of NTCIR-12 Lifelog Task. In *Proceedings of the 12th NTCIR Conference on Evaluation of Information Access Technologies, National Center of Sciences, Tokyo, Japan, June 7-10, 2016*, Noriko Kando, Tetsuya Sakai, and Mark Sanderson (Eds.). National Institute of Informatics (NII). http://research.nii.ac.jp/ntcir/workshop/OnlineProceedings12/pdf/ntcir/OVERVIEW/01-NTCIR12-OV-LIFELOG-GurrinC.pdf

[5] Cathal Gurrin, Hideo Joho, Frank Hopfgartner, Liting Zhou, Duc-Tien Dang-Nguyen, Rashmi Gupta, and Rami Albatal. 2017. Overview of NTCIR-13 Lifelog-2 Task. In *The 13th NTCIR Conference, Evaluation of Information Access Technologies, National Center of Sciences, Tokyo, Japan, December 5-8, 2017.* National Institute of Informatics (NII). https://research.nii.ac.jp/ntcir/workshop/OnlineProceedings13/pdf/ntcir/01-NTCIR13-OV-LIFELOG-GurrinC.pdf

[6] 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.

[7] Cathal Gurrin, Alan F. Smeaton, and Aiden R. Doherty. 2014. LifeLogging: Personal Big Data. *Found. Trends Inf. Retr.* 8, 1 (jun 2014), 1–125. https://doi.org/10.1561/1500000033

[8] Zhiyu He, Jiayu Li, Wenjing Wu, Min Zhang, Yiqun Liu, and Shaoping Ma. 2022. THUIR-LL at the NTCIR-16 Lifelog-4 Task: Enhanced Interactive Lifelog Search Engine. (2022).

[9] Junnan Li, Dongxu Li, Silvio Savarese, and Steven Hoi. 2023. Blip-2: Bootstrapping language-image pre-training with frozen image encoders and large language models. *arXiv preprint arXiv:2301.12597* (2023).

[10] Junnan Li, Dongxu Li, Caiming Xiong, and Steven Hoi. 2022. Blip: Bootstrapping language-image pre-training for unified vision-language understanding and generation. In *International Conference on Machine Learning.* PMLR, 12888–12900.

[11] Junnan Li, Ramprasaath Selvaraju, Akhilesh Gotmare, Shafiq Joty, Caiming Xiong, and Steven Chu Hong Hoi. 2021. Align before fuse: Vision and language representation learning with momentum distillation. *Advances in neural information processing systems* 34 (2021), 9694–9705.

[12] Thao-Nhu Nguyen, Tu-Khiem Le, Van-Tu Ninh, Ly-Duyen Tran, Manh-Duy Nguyen, Minh-Triet Tran, Binh T Nguyen, Annalina Caputo, Cathal Gurrin, Graham Healy, et al. 2022. DCU and HCMUS at NTCIR-16 Lifelog-4. In *Proceedings of the 16th NTCIR Conference on Evaluation of Information Access Technologies*. NTCIR.

[13] Thao-Nhu Nguyen, Tu-Khiem Le, Van-Tu Ninh, Minh-Triet Tran, Thanh Binh Nguyen, Graham Healy, Sinéad Smyth, Annalina Caputo, and Cathal Gurrin. 2022. LifeSeeker 4.0: An Interactive Lifelog Search Engine for LSC'22. In *Proceedings of the 5th Annual on Lifelog Search Challenge* (Newark, NJ, USA) *(LSC '22)*. Association for Computing Machinery, New York, NY, USA, 14–19. https://doi.org/10.1145/3512729.3533014

[14] Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, Gretchen Krueger, and Ilya Sutskever. 2021. Learning Transferable Visual Models From Natural Language Supervision. arXiv:2103.00020 [cs.CV]

[15] Ly Duyen Tran, Binh Nguyen, Liting Zhou, and Cathal Gurrin. 2023. MyEachtra: Event-Based Interactive Lifelog Retrieval System for LSC'23. In *Proceedings of the 6th Annual ACM Lifelog Search Challenge* (Thessaloniki, Greece) *(LSC '23)*. Association for Computing Machinery, New York, NY, USA, 24–29. https://doi.org/10.1145/3592573.3593100

[16] 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* (Taipei, Taiwan) *(LSC '21)*. Association for Computing Machinery, New York, NY, USA, 11–16. https://doi.org/10.1145/3463948.3469064

[17] Ly-Duyen Tran, Dongyun Nie, Liting Zhou, Binh Nguyen, and Cathal Gurrin. 2023. VAISL: Visual-Aware Identification of Semantic Locations in Lifelog. In *International Conference on Multimedia Modeling*. Springer, 659–670.

[18] Quang-Linh Tran, Ly-Duyen Tran, Binh Nguyen, and Cathal Gurrin. 2023. MemoriEase: An Interactive Lifelog Retrieval System for LSC'23. In *Proceedings of the 6th Annual ACM Lifelog Search Challenge* (Thessaloniki, Greece) *(LSC '23)*. Association for Computing Machinery, New York, NY, USA, 30–35. https://doi.org/10.1145/3592573.3593101

[19] Sam Witteveen and Martin Andrews. 2019. Paraphrasing with large language models. *arXiv preprint arXiv:1911.09661* (2019).

[20] Liting Zhou, Cathal Gurrin, Duc-Tien Dang-Nguyen, Graham Healy, ChenYang Lyu, Tianbo Ji, Longyue Wang, Joho Hideo, Ly-Duyen Tran, and Alam Naushad. 2023. Overview of the NTCIR-17 Lifelog-5 Task. In *Proceedings of the 17th NTCIR Conference on Evaluation of Information Access Technologies*. National Institute of Informatics.

[21] L. Zhou, C. Gurrin, G. Healy, H. Joho, B. Nguyen, R. Albatal, F. Hopfgartner, and D.-T. Dang-Nguyen. 2022. Overview of the NTCIR-16 Lifelog-4 Task, In 16th NTCIR Conference : Evaluation of Information Access Technologies, M.P. Kato, T. Yamamoto, and Z. Dou (Eds.). *Proceedings of the 16th NTCIR Conference on Evaluation of Information Access Technologies*, 130–135. https://eprints.whiterose.ac.uk/188351/ © 2022 National Institute of Informatics..

APPENDIX

In this appendix, we provide the table of challenging queries and the search interface for some difficult queries.

Figure 4: Search Interface for query 17002: Taking Medication.

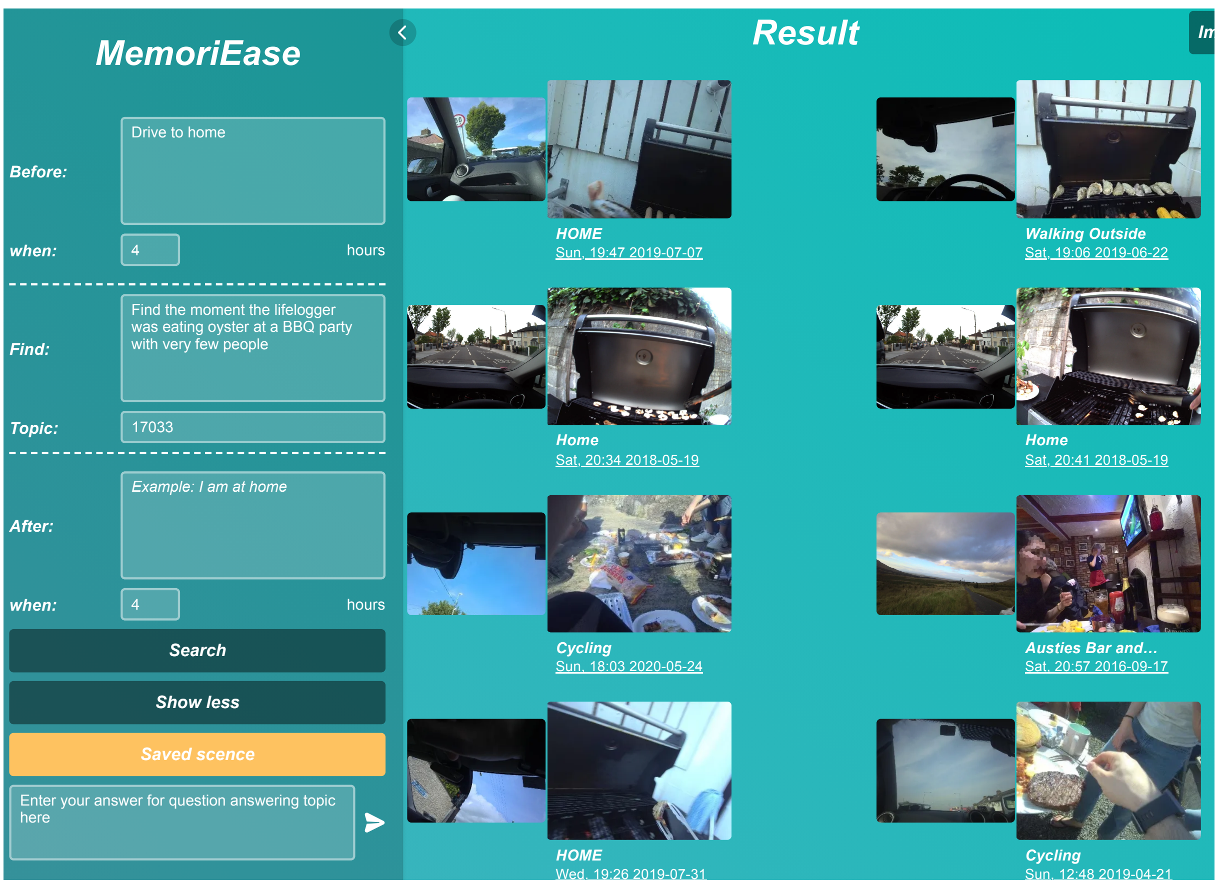

Figure 5: Search Interface for query 17033: Oyster.

Topic ID	Type	Title	Description	Narrative	Challenge
17002	adhoc	Taking Medication	Find examples of when I was taking medication or vitamins at home	To ensure relevance, the images must clearly display either medication or vitamins being utiliz	medication or vitamins is small and can be hardly to recognise in the images. In addition, the lifelogger should take that medication so the medication or vitamins should be on the hand of the lifelogger before of taking them
17013	adhoc	Listening Music	Find examples of when I was listening to music by Elton John on an airplane	To be considered relevant, the moment should occur exclusively while the lifelogger is aboard an airplane. These images should capture me relishing Elton John's songs within the airplane's cabin.	This query requires a metadata of listening to music to find the moment of the lifelogger listening to music on airplane by Elton John. Another approach is to find the image with Elton John on the display in the airplane, but this can be challenging for the BLIP-2 model.
17018	known item	'Get back' on the roof	Find any moments in which the lifelogger was watching the Beatles rooftop concert on tv.	The lifelogger was watching the Beatles rooftop concert on tv (not in music metadata). The lifelogger was at home watching youtube on TV for about 90 minutes, after doing some computer work at home.	This query requires a moment of watching TV on a music concert of a band. However, this can be difficult to find the moment because there is no metadata for this.
17033	known item	Oyster	Find the moment the lifelogger was eating oyster	It was at home and there were a few people at the BBQ. The lifelogger had driven home before cooking lots of different foods on the BBQ.	For the BBQ party, there is a lot of food on the table and oysters can be difficult to recognise in the image.
17037	known item	Having lunch with Dermot	Find the moment when the lifelogger was having lunch with Dermot	Dermot was a guest speaker at my lecture. After lunch, he gave a lecture to my class about Lessons in Innovation and Entrepreneurship while I was sitting in the front row.	The query is challenge because it requires a specific moment of having lunch with Dermot but this name can challenge the BLIP-2 model because there is no prior knowledge about him.

Table 3: Challenging queries of NTCIR-17 Lifelog5 for the MemoriEase.