NTCIR-18: Proceedings of the 18th NTCIR Conference on Evaluation ofInformation Access Technologies, June 10-13, 2025, Tokyo, Japan
DOI: https://doi.org/10.20736/0002002048

DCU MemoriEase at the NTCIR-18 Lifelog 6 Task

Quang-Linh Tran
ADAPT Centre, School of Computing,
Dublin City University
Dublin 9, Ireland
quang-linh.tran2@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 participation of the MemoriEase lifelog retrieval system in the NTCIR-18 Lifelog 6 Task. This current MemoriEase system is an automatic and enhanced version of the MemoriEase system at the Lifelog Search Challenge 2024 (LSC). We report our methods for the two core sub-tasks in the NTCIR-18 Lifelog 6 task, Lifelog Semantic Access (LSAT) and Lifelog Question Answer (LQAT). We enhance the main architecture of the MemoriEase system utilizing the BLIP2 and CLIP embedding models to extract visual embedding and perform a comparison between the two models. In addition, we also use pseudo-relevance feedback for ad-hoc queries. For the LQAT sub-task, we use our retrieval model as the retriever and GPT-4o as a reader to generate answers to questions. Results of the LSAT sub-task show that our system found 369 images in 1,995 relevant images. The performance on known-item search queries is higher than on Ad-hoc queries, with 28.22% R@5 compared to 5.98% R@5, respectively. In the LQAT sub-task, the LLM model generates 8 correct answers in 24 questions. Although the performance is not high, it shows the advantages and drawbacks of the MemoriEase retrieval system and the QA model.

Keywords

lifelog, information retrieval, personal data

Team Name

MEaseDCU

Subtasks

Lifelog Semantic Access - LSAT
Lifelog Question Answer - LQAT

1 Introduction

Lifelogging is a process of automatically collecting personal daily life data for further use [7]. Lifelog data includes multiple modalities, including visual (images/videos), textual (descriptions), tabular (metadata), etc. After collection, this data can be stored in a secure storage and analyzed for use in multiple applications. For example, lifelog data can be used in health monitoring [4], lifestyle analytics [12], and memory enhancement [5]. Developing methods for retrieval from lifelogs is an active area of research with several benchmarks, one of which is the Lifelog task at NTCIR is one of them. The lifelog retrieval task is a text-to-image retrieval task on lifelog data. In this task, textual queries are provided, requiring the relevant lifelog images to be identified for each query. Depending on the type of query, the definition of relevant images may be different. For known-item queries, a single event in the lifelog data is considered as the answer, where an event is a sequence of images at a specific time depicting a single activity of the lifelogger, such as eating or walking. Finding one image in the event is satisfactory for solving the task. This type of query aims to enhance the memory of life loggers when they want to reminisce about an event in the past. Another type of query is ad-hoc queries, which involve finding multiple events depicting the same activity but at different times and locations. For example, “Finding all the time I work out” is an ad-hoc query requiring finding all the events of working out. This type of query helps lifeloggers to have a view of their lifestyle and habits. These queries belong to the LSAT sub-task in the NTCIR-18 Lifelog 6 task [31]. For the LQAT sub-task, a question is posed to find an answer based on the lifelog data, such as “When did I last walk along the beach?”. This task promotes a natural ask-and-answer manner when interacting with lifelog data, and from the question, lifeloggers can acquire a significant number of insights into their lifelog. In the NTCIR-18 Lifelog 6 [31], the MEaseDCU team participated in both sub-tasks described above.

The NII Testbeds and Community for Information Access Research (NTCIR) is a well-known place for organizing evaluations of tasks related to information access. The Lifelog task is one of the core tasks and has been organized for six iterations. The MEaseDCU team has participated in this benchmark challenge for the second time. In this NTCIR-18 Lifelog 6 task [31], we present an enhanced version of the MemoriEase system with several improvements in retrieval engines for the LSAT sub-task. In addition, this is also the first time we have participated in the LQAT sub-task, which involved integrating a Large Language Model (LLM) into the retrieval engine for question-answering.

For our participation in the NTCIR-18 Lifelog 6 task, we have adapted and enhanced our existing interactive MemoriEase lifelog retrieval system [27] to enable automatic operation. There are four main new features in this version. Firstly, we experiment with two state-of-the-art (SOTA) embedding models, BLIP2 [14] and CLIP [19], to evaluate their robustness in lifelog retrieval tasks. Secondly, we employ a pseudo-relevance feedback technique [29] for ad-hoc queries to enhance the performance through visual similarity

search. Thirdly, we use ChatGPT[1] to create a concrete and concise query from the original query to experiment with the effect of the query on retrieval. Finally, we utilize a state-of-the-art LLM [11] for the LQAT task, which uses retrieved data from the retrieval engine to generate answers. We believe that with these improvements, MemoriEase can achieve high performance in the NTCIR Lifelog 6 task.

2 Related Work

This section provides an overview of related work in the field of lifelog retrieval and question answering. The lifelogging concept dates back to 1945 when Vannevar Bush proposed Memex as an external human memory [3]. However, this field of research did not develop until recently, when the availability of low-cost wearable sensors and data storage became more accessible. The first test collection for lifelog research was developed for NTCIR 12 Lifelog task [6]. From that time, more benchmark datasets and retrieval methods have been introduced. LSC [8] and NTCIR Lifelog task [30] are the two well-known benchmark challenges for evaluating lifelog retrieval systems in both interactive and automatic manners. Meanwhile, state-of-the-art lifelog retrieval systems have evolved from keyword-based retrieval [15] to embedding-based retrieval [23, 24], aligned with general text-to-image retrieval. In the last lifelog task at NTCIR 17 [30], three systems participated in LSAT automatic runs, and two systems participated in LSAT interactive runs. The DCU MemoriEase [26] team developed an automatic retrieval system using the BLIP-2 embedding model [14], enhanced by Elasticsearch[2] indexing. They integrated ChatGPT[3] for preprocessing and post-processing, achieving an mAP score of 0.2713 and a P@5 score of 0.3707, demonstrating effective and practical retrieval performance. The HCMUS-DCU LifeInsight [17] team participated in the automatic runs using Semantic Role Labeling (SRL) and LLMs to extract entities, identify temporal events, and generate context-aware prompts for lifelog retrieval. They leveraged vector databases and visual-language models to improve search performance, emphasizing subqueries and contextual analysis. Their approach achieved a mAP of 0.2924 and p@5 of 0.4098, outperforming DCU MemoriEase and DCU Memento. The DCU-Memento [1] system [30] utilized CLIP models [19] for image-text retrieval through a multi-stage ranking process. They submitted nine runs, using individual and ensemble embeddings, achieving a score of 0.1734 with 450 relevant items. The CLIP ViT-G/14 model showed high recall, while the ensemble of CLIP ViT-G/14 and ResNet50x64 achieved the highest precision and strong performance in MRR and NDCG. On the interactive runs, the HCMUS-DCU LifeInsight [28] system showed expert users performed better, though some novices achieved good results, user feedback highlighted the system's efficiency and engagement. The best run, HCMS-INTERACTIVE-07, achieved a mAP score of 0.1686 across 41 topics, with noted areas for improvement in usability and technological advancement. The UA Memoria [20] team developed an interactive lifelog retrieval system using YOLOv7, GRiT, OCR, Places365, and ClipCap for image annotation and retrieval. They focused on hierarchical event segmentation based on time, location, and image similarity to enhance memory organization. Their submission achieved the highest mAP scores of 0.5968 for known-item tasks and 0.2895 for ad-hoc tasks, outperforming other interactive entries.

The Lifelog Search Challenge 2024 (LSC'24) attracted several innovative retrieval systems aimed at enhancing lifelog retrieval performance. MemoriEase 2.0 [27] built on its predecessor by integrating conversational search, visual similarity search, and retrieval-augmented generation, achieving a 40% Recall@1 on initial hints and successfully solving 8 out of 10 topics in evaluations. Eagle team [16] focused on bridging the gap between expert and novice users by incorporating implicit user interactions, such as eye movements, to optimize search result displays and improve user experience through an automated search flow. SnapSeek [10] introduced a semantic search system leveraging CLIP [19], BLIP-2 [14], BEIT-3, and all-mpnet-base-v2 models, combined with Milvus [4] for efficient indexing, enabling context-aware search through metadata enhancement, query auto-parsing, and an improved ranking system. VISIONE [2], a system initially designed for video retrieval, was adapted to lifelogging by organizing images temporally, though it relies solely on visual content analysis without indexing metadata like GPS or local time. LifeXplore [18] refined free-text search using FAISS [5] and CLIP [19] while integrating flexible multimodal queries, a new system architecture, and an optimized GUI with advanced filtering options. These systems collectively pushed the boundaries of lifelog retrieval by improving search accuracy, user interaction, and multimodal integration, making lifelogging more accessible and practical for users. LifeSeeker 6.0 [13] introduced an improved user interface and backend by integrating contrastive learning between text inputs and images, enhancing retrieval accuracy and efficiency. Libro [9] adapted video search techniques from the Vibro system [22] by treating lifelog data as continuous video frames, incorporating text-to-image, image-to-image search, metadata filtering, and a graph-based dataset visualization for better exploration. LifeGraph 4 [21] leveraged multimodal knowledge graphs with event-based clustering based on temporal and spatial relations, enriched with Vision-Language Model (VLM) descriptions to improve contextual understanding. MyEachtraX [25] took a mobile-first approach, integrating LLM and Multimodal LLM (MLLM) to enhance query parsing, post-processing, and question-answering in lifelog retrieval, achieving 72.2% accuracy in evaluations, though identifying retrieval as a bottleneck for future improvements. From the approaches of these retrieval systems, we aim to enhance our system by incorporating CLIP embedding models, which are widely used by other systems. In addition, with the development of LLM in various domains, we also use it as a reader for the lifelog QA sub-task.

3 MemoriEase System

This section describes the MemoriEase system, from data processing and indexing to retrieval. In addition, we will provide more information about the improvements made to this version. Figure 1 illustrates the overall architecture of the MemoriEase system.

[1]https://chatgpt.com/
[2]https://elastic.co/
[3]https://chatgpt.com/
[4]https://milvus.io/
[5]https://faiss.ai/

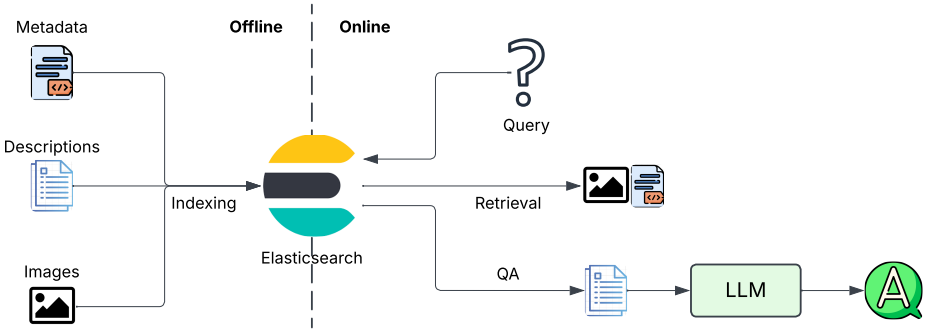

Figure 1: MemoriEase system architecture

3.1 Data Processing and Indexing

The lifelog dataset used for this task includes 18 months of lifelog data from one active lifelogger. The core image dataset has over 700.000 lifelog images at a resolution of 1024x768 pixels captured by a Narrative Clip device. In addition, the associated metadata includes information such as time, location, etc. There is also a visual concept dataset, which is extracted from the core image dataset, depicting the information of the scene in the images. Furthermore, we also used the descriptions of lifelog images from previous work to enhance the dataset.

From the lifelog images and metadata, we extract the embedding of images from the BLIP2 and CLIP models. The embedding size of BLIP2 is 256, while the embedding size of CLIP is 1024, which is 4 times higher. In addition, we extract several pieces of information from metadata such as city, local time, hour, day of the week, weekend, and time period to play the role of filters. Along with the descriptions, we index them into the Elasticsearch vector database.

3.2 Retrieving Events

In the retrieval stage, the query is first processed by several rules before being converted into embedding vectors by the two embedding models. These rules include uncapitalizing and removing filter-related information. We will compare the performance of two embedding models in the same setting to see which model encodes images and queries better (BLIP2 vs CLIP). The Elasticsearch database then calculates the cosine similarity of query embeddings and image embeddings to retrieve the top K-ranked list of images. If there are filters associated with the query, they are automatically extracted. These filters are related to metadata in the dataset, such as time, location, and date. For example, in the query "How many days did I cycle in April 2020?" the system extracts the local time filter as 01-04-2020 to 30-04-2020. The ranked list of images filters out the images with metadata that does not match the filter.

We also experiment with using a set of LLM-generated queries from original queries. Because the provided queries contain a lot of information in a long sentence (from title, description to narrative), the embedding model is distracted from focusing on key information. We prompt ChatGPT with the following prompt: "You are a helpful text-to-image search assistant. Can you please rewrite these filters to make it concise, focus on main points, suitable for CLIP and BLIP2 model to search: Queries." to generate a new set of queries. These queries are processed using the same procedure as the original queries. These queries are denoted CQ (concise query) instead of Q (original queries) in the experiment. We will compare the performance of the BLIP2 model in the same settings except for query input to see the difference in the performance of the two query sets.

For the ad-hoc queries, we implement a pseudo-relevance feedback technique to improve the accuracy of searching. Specifically, we first use the query to retrieve and get the top 3 images. We then calculate the average of the vector embeddings of these 3 images along with the embedding of the original query and perform another retrieval round. This step captures the visual information of correct images and finds similar ones. The result is compared with the textual query ranked list to see which method performs better. We denote experiments with pseudo-relevance feedback as RF while others are as NoRF.

3.3 Question-Answering

To solve the LQAT sub-task, we enhance the retrieval system by incorporating an LLM model to answer questions. The question is classified into two types: visual-related questions and metadata-related questions. We use a rule-based approach based on question words to classify. This approach classifies the questions starting with "What, Who, Why" as visual-related questions and the others as metadata-related questions. Depending on the type of question, we employ two different approaches to find the answer to the question.

For visual-related questions, we use the query from the question to retrieve the top 10 images and prompt the LLM model to generate an answer. For example, to resolve the question "What is the brand of the drum set in my sister's house?", we first use the query "the brand of the drum set in my sister's house" to retrieve the top 10 images. We use the prompt "You are a helpful visual question-answering assistant. You are provided with several images and a question. Reasoning yourself to choose the correct image that provides an answer to the question and gives me a textual answer. Provide a short answer with no explanation. Question: What is the brand of the drum set in my sister's house? Images: </images>. Answer: ". The final output is LLM's response to the question. Figure 2 illustrates the result of the question in the user interface of the MemoriEase system.

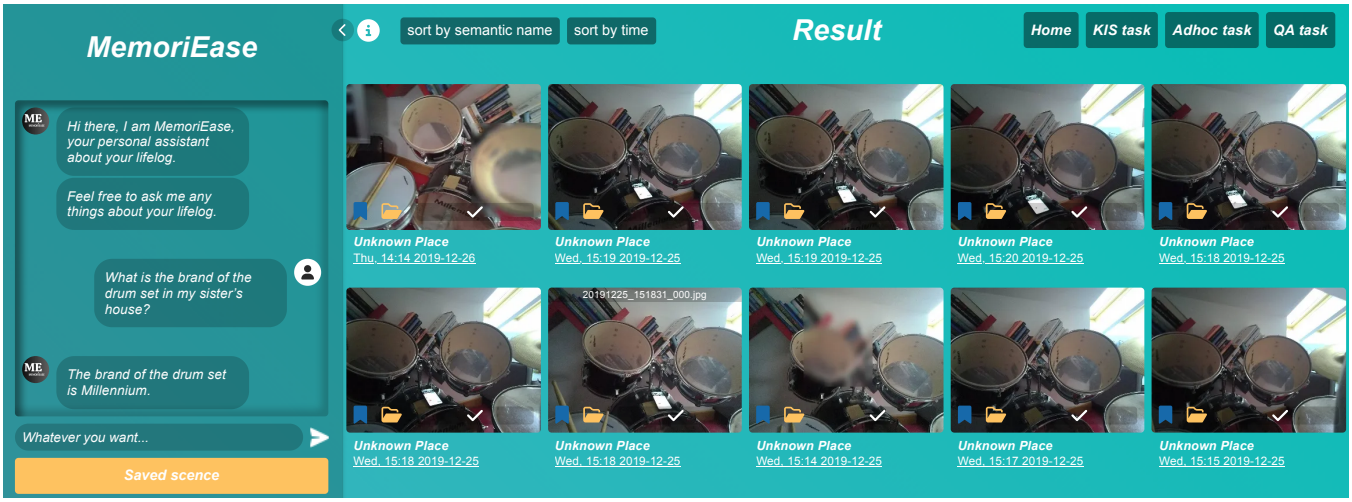

Figure 2: An example of QA queries and model's response.

For metadata-related questions, the retrieval system retrieves the top 100 descriptions and reranks them to a short list of the top 10 most relevant descriptions. The reranking model [6] is a cross-encoder reranking model fine-tuned on lifelog data. In addition to the top 10 descriptions of events in lifelog data, we also append it with the top 10 most relevant images to the question to create a list of 20 visual and textual items. This aims to enhance the list of contexts to incorporate both textual descriptions and visual images. We then feed the LLM with the prompt: "You are a helpful assistant who can answer the question based on the provided context. Reasoning over the provided data to find the correct information for the question. Provide a short answer with no explanation. Question: <question>. Contexts: <contexts>. Images: <images>." The generated answer will be used for the final submission.

4 Experiment

This section discusses the topics from the organizers of the NTCIR-18 Lifelog 6 task, the results of different runs of our system, and performance and error analysis.

4.1 Topics

There are 13 topics for KIS queries, 13 topics for Ad-hoc queries and 24 topics for the LQAT sub-task. Each topic has an ID, title, description, and narrative. The title is a short summary of the query, such as "Trains in the attic". The description provides more detailed information, such as "Looking at model trains in somebody's attic room.". However, there is more constraint information in the narrative, such as "I was in an attic looking at trains on shelves in an attic. There was someone else there. It was not in my home. Any examples of attics in my own home are not considered relevant.". We use both description and narrative to generate concise queries using ChatGPT. In the LQAT sub-task, questions usually ask about the information in the images, such as "I had a healthy birthday celebration with a lot of fruit and a birthday cake. List the fruits that were on the table.". The answer can be found in the images by analyzing the birthday party images. However, questions like "How many days did I cycle in April 2020?" need the metadata of cycling images to find the answers. In previous sections, we classify the questions and propose different approaches for different question types.

4.2 Results

We analyze the results of the LSAT sub-task of two types of queries to have a deeper view of the system's performance. Table 1 provides the information of performances of the system in different metrics for Ad-hoc queries and KIS queries. For the ad-hoc queries, the BLIP-Q-RF combination finds the most correct relevant images, with 286 images in 1746 relevant images. However, the best combination is BLIP-CQ-NoRF, which dominates most of the metrics. It achieves 18% MAP, 68.06% MRR and 44.06% R@100. We expect the relevance feedback (RF) will achieve good performance in this query, and it actually finds the most correct images, but the rank of the correct images is not high, so it decreases the other metrics. The CLIP embedding model is less effective for this type of query than the BLIP2 embedding model, and the concise, ChatGPT summarized queries are better than original queries in Ad-hoc queries. Figure 3 illustrates the interpolated precision at different recall levels of four combinations in the Ad-hoc queries. There are two distinct groups in the performance, which are BLIP with NoRF and CLIP & BLIP - RF.

Table 1b shows the system's performance in the KIS task. There are 249 relevant images for 13 queries, and our best system found 83 correct images. The best combination is BLIP-Q-NoRF with 52.42% MRR, 28.22% R@5, and 63.49% R@100. The performance of BLIP-Q-NoRF and BLIP-Q-RF is the same because relevance feedback is not used in KIS queries. The CLIP model finds more correct images, but the BLIP model achieves a better performance because of the rank of correct images of CLIP. The concise queries do not enhance the performance of KIS queries compared to original queries. This indicates that KIS queries need more detailed information to find

[6] https://huggingface.co/cross-encoder/ms-marco-MiniLM-L6-v2

Table 1: Performance of different approaches in Ad-hoc and KIS queries. CLIP/BLIP: embedding models, Q/CQ: original queries vs concise, LLM summarized queries, NoRF/RF: No relevance feedback vs pseudo relevance feedback

(a) Ad-hoc Queries

Ad-hoc Queries	MAP	MRR	NDCG	R@5	R@20	R@100	P@5	P@20	P@100	#relevance	#retrieved_relevance
CLIP - Q - NoRF	0.0673	0.4391	0.1823	0.015	0.1223	0.2426	0.2615	0.2462	0.21	1746	273
BLIP - Q - RF	0.0754	0.2827	0.1398	0.0136	0.0382	0.132	0.2462	0.2385	0.22	1746	**286**
BLIP - Q - NoRF	0.1661	0.4616	0.3058	0.055	0.1518	0.3802	0.3692	0.2923	0.1815	1746	236
BLIP- CQ - NoRF	**0.1805**	**0.6806**	**0.3528**	**0.0598**	**0.167**	**0.4406**	**0.4308**	**0.3423**	**0.2192**	1746	285

(b) KIS Queries

KIS Queries	MAP	MRR	NDCG	R@5	R@20	R@100	P@5	P@20	P@100	#relevance	#retrieved_relevance
CLIP - Q - NoRF	0.2	0.4274	0.3279	0.191	0.4357	0.494	0.2	0.1346	**0.0638**	249	**83**
BLIP - Q - RF	0.2702	0.5242	0.4287	0.2822	0.5035	0.6349	0.2462	0.1462	0.0554	249	72
BLIP - Q - NoRF	0.2702	**0.5242**	**0.4287**	**0.2822**	**0.5035**	**0.6349**	**0.2462**	**0.1462**	0.0554	249	72
BLIP- CQ - NoRF	**0.2709**	0.3425	0.3785	0.2698	0.3847	0.5348	0.2	0.1192	0.0554	249	72

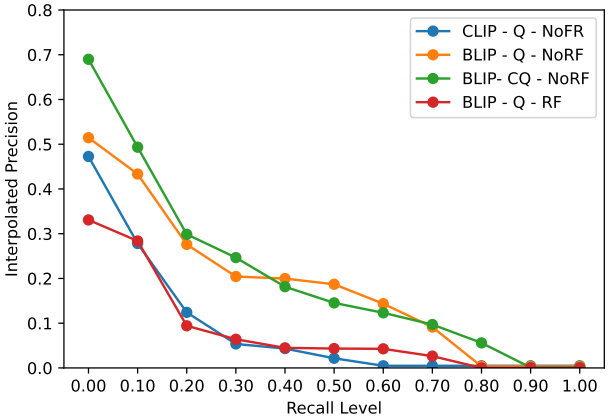

Figure 3: Ad-hoc queries performance

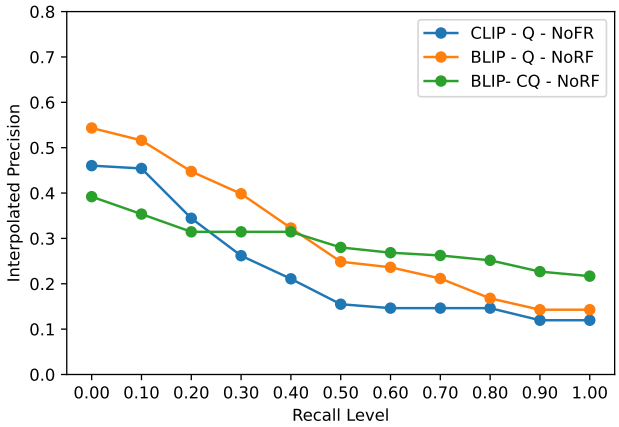

Figure 4: KIS queries performance

the correct images. Figure 4 illustrates the interpolated precision at different recall levels of three combinations. The chart shows that the BLIP model (both BLIP-Q and BLIP-CQ) outperforms the CLIP model across all recall levels. Among the query types in the BLIP model, BLIP-Q has higher precision at lower recall levels, while BLIP-CQ maintains more stable performance across different recall levels.

Table 2 shows the questions, ground-truth answers and predicted answers from our system. Overall, we predict 8 correct answers in 24 questions. Our correct answers are mainly from the visual-related questions that the answer in the lifelog images, such as "What is the brand of the drum set in my sister's house?". With 6 correct answers from visual-related questions and 2 answers from metadata-related questions, it shows the drawback of our system for metadata-related questions and the challenge of that type of question. We do not have the ground-truth images to check if our retrieved images are correct, but the main reason for wrong answers in metadata-related images is the lack of or the wrong input images/descriptions to the LLM models. There are 5 questions that the LLM model cannot determine the answer because there is no information in the prompt. This highlights the importance of the retrieval model in the lifelog QA task. There is a lot of room for improvement in this task, and we will enhance our retrieval and QA models to improve performance.

5 Conclusions

In this paper, we presented the MemoriEase lifelog retrieval system that participated in the NTCIR-18 Lifelog 6 Task. Our system introduced multiple enhancements over the previous LSC 2024 version, including the integration of two SOTA embedding models, pseudo-relevance feedback for ad-hoc queries, and LLM for the LQAT sub-task. The evaluation results demonstrated that MemoriEase was able to retrieve 369 relevant images out of 1,995 images in the LSAT task. In the LQAT task, our system correctly answered 8 out of 24 questions, with strong performance in visual-based queries but limitations in metadata-related questions. These results highlight both the potential and the current challenges of using

retrieval-augmented LLMs for lifelog-based question answering. For future work, we aim to refine our LQAT approach by improving retrieval accuracy and providing more structured input to LLMs to minimize incorrect answers due to incomplete context.

Acknowledgement

This research was conducted with the financial support of Research Ireland at ADAPT, the Research Ireland Centre for AI-Driven Digital Content Technology at Dublin City University [13/RC/21-06_P2]. For the purpose of Open Access, the author has applied a CC BY public copyright licence to any Author Accepted Manuscript version arising from this submission.

References

[1] Naushad Alam, Yvette Graham, and Cathal Gurrin. 2023. DCU at the NTCIR-17 Lifelog-5 Task. *NII Institutional Repository* (2023). doi:records/2001337

[2] Giuseppe Amato, Paolo Bolettieri, Fabio Carrara, Fabrizio Falchi, Claudio Gennaro, Nicola Messina, Lucia Vadicamo, and Claudio Vairo. 2024. Will VISIONE Remain Competitive in Lifelog Image Search?. In *Proceedings of the 7th Annual ACM Workshop on the Lifelog Search Challenge* (Phuket, Thailand) *(LSC '24)*. Association for Computing Machinery, New York, NY, USA, 58–63. doi:10.1145/3643489.3661122

[3] Vannevar Bush. 1996. As we may think. *Interactions* 3, 2 (March 1996), 35–46. doi:10.1145/227181.227186

[4] Junho Choi, Chang Choi, Hoon Ko, and Pankoo Kim. 2016. Intelligent healthcare service using health lifelog analysis. *Journal of medical systems* 40 (2016), 1–10.

[5] Tilman Dingler, Passant El Agroudy, Rufat Rzayev, Lars Lischke, Tonja Machulla, and Albrecht Schmidt. 2021. Memory augmentation through lifelogging: opportunities and challenges. *Technology-augmented perception and cognition* (2021), 47–69.

[6] Cathal Gurrin, Hideo Joho, Frank Hopfgartner, Liting Zhou, and Rami Albatal. 2016. Ntcir lifelog: The first test collection for lifelog research. In *Proceedings of the 39th International ACM SIGIR conference on Research and Development in Information Retrieval*. 705–708.

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

[8] Cathal Gurrin, Liting Zhou, Graham Healy, Werner Bailer, Duc-Tien Dang Nguyen, Steve Hodges, Björn Þór Jónsson, Jakub Lokoč, Luca Rossetto, Minh-Triet Tran, et al. 2024. Introduction to the seventh annual lifelog search challenge, lsc'24. In *Proceedings of the 2024 International Conference on Multimedia Retrieval*. 1334–1335.

[9] Nico Hezel, Konstantin Schall, Bruno Schilling, Klaus Jung, and Kai Uwe Barthel. 2024. Libro - Lifelog Search Browser. In *Proceedings of the 7th Annual ACM Workshop on the Lifelog Search Challenge* (Phuket, Thailand) *(LSC '24)*. Association for Computing Machinery, New York, NY, USA, 70–75. doi:10.1145/3643489.3661124

[10] Minh-Quan Ho-Le, Huy-Hoang Do-Huu, Duy-Khang Ho, Nhut-Thanh Le-Hinh, Hoa-Vien Vo-Hoang, Van-Tu Ninh, and Minh-Triet Tran. 2024. SnapSeek: An Interactive Lifelog Acquisition System for LSC'24. In *Proceedings of the 7th Annual ACM Workshop on the Lifelog Search Challenge* (Phuket, Thailand) *(LSC '24)*. Association for Computing Machinery, New York, NY, USA, 24–29. doi:10.1145/3643489.3661116

[11] Aaron Hurst, Adam Lerer, Adam P Goucher, Adam Perelman, Aditya Ramesh, Aidan Clark, AJ Ostrow, Akila Welihinda, Alan Hayes, Alec Radford, et al. 2024. Gpt-4o system card. *arXiv preprint arXiv:2410.21276* (2024).

[12] Gunjan Kumar, Houssem Jerbi, Cathal Gurrin, and Michael P O'Mahony. 2014. Towards activity recommendation from lifelogs. In *Proceedings of the 16th international conference on information integration and web-based applications & services*. 87–96.

[13] Hoang-Bao Le, Thao-Nhu Nguyen, Tu-Khiem Le, Minh-Triet Tran, Thanh-Binh Nguyen, Van-Tu Ninh, Liting Zhou, and Cathal Gurrin. 2024. LifeSeeker 6.0: Leveraging the linguistic aspect of the lifelog system in LSC'24. In *Proceedings of the 7th Annual ACM Workshop on the Lifelog Search Challenge* (Phuket, Thailand) *(LSC '24)*. Association for Computing Machinery, New York, NY, USA, 53–57. doi:10.1145/3643489.3661121

[14] 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. In *International conference on machine learning*. PMLR, 19730–19742.

[15] Thao-Nhu Nguyen, Tu-Khiem Le, Van-Tu Ninh, Annalina Caputo, Graham Healy, Sinéad Smyth, Minh-Triet Tran, and Nguyen Thanh Binh. 2023. LifeSeeker: an interactive concept-based retrieval system for lifelog data. *Multimedia Tools and Applications* 82, 24 (2023), 37855–37876.

[16] Thang-Long Nguyen-Ho, Onanong Kongmeesub, Minh-Triet Tran, Dongyun Nie, Graham Healy, and Cathal Gurrin. 2024. EAGLE: Eyegaze-Assisted Guidance and Learning Evaluation for Lifeloging Retrieval. In *Proceedings of the 7th Annual ACM Workshop on the Lifelog Search Challenge* (Phuket, Thailand) *(LSC '24)*. Association for Computing Machinery, New York, NY, USA, 18–23. doi:10.1145/3643489.3661115

[17] Thang-Long Nguyen-Ho, Gia Huy Vuong, Van-Son Ho, Tien-Thanh Nguyen-Dang, Xuan-Dang Thai, Minh-Khoi Pham, Tu-Khiem Le, Van-Tu Ninh, and Minh-Triet Tran. 2023. Automatic Sub-Task Focus: LifeInsight's Contribution to NTCIR-17 Lifelog-5. *NII Institutional Repository* (2023). doi:records/2001334

[18] Martin Rader and Klaus Schoeffmann. 2024. lifeXplore at the Lifelog Search Challenge 2024. In *Proceedings of the 7th Annual ACM Workshop on the Lifelog Search Challenge* (Phuket, Thailand) *(LSC '24)*. Association for Computing Machinery, New York, NY, USA, 64–69. doi:10.1145/3643489.3661123

[19] 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] https://arxiv.org/abs/2103.00020

[20] Ricardo Ribeiro, Alexandre Gago, Bernardo Kaluza, Josefa Pandeirada, and António Neves. 2023. MEMORIA: A Memory Enhancement and MOment RetrIeval Application at the NTCIR-17 Lifelog-5 Task. *NII Institutional Repository* (2023). doi:10.20736/0002001335

[21] Luca Rossetto, Athina Kyriakou, Svenja Lange, Florian Ruosch, Ruijie Wang, Kathrin Wardatzky, and Abraham Bernstein. 2024. LifeGraph 4 - Lifelog Retrieval using Multimodal Knowledge Graphs and Vision-Language Models. In *Proceedings of the 7th Annual ACM Workshop on the Lifelog Search Challenge* (Phuket, Thailand) *(LSC '24)*. Association for Computing Machinery, New York, NY, USA, 88–92. doi:10.1145/3643489.3661127

[22] Konstantin Schall, Nico Hezel, Klaus Jung, and Kai Uwe Barthel. 2023. Vibro: Video Browsing with Semantic and Visual Image Embeddings. In *MultiMedia Modeling: 29th International Conference, MMM 2023, Bergen, Norway, January 9–12, 2023, Proceedings, Part I* (Bergen, Norway). Springer-Verlag, Berlin, Heidelberg, 665–670. doi:10.1007/978-3-031-27077-2_56

[23] Ly-Duyen Tran, Manh-Duy Nguyen, Duc-Tien Dang-Nguyen, Silvan Heller, Florian Spiess, Jakub Lokoč, Ladislav Peška, Thao-Nhu Nguyen, Omar Shahbaz Khan, Aaron Duane, Björn þór Jónsson, Luca Rossetto, An-Zi Yen, Ahmed Alateeq, Naushad Alam, Minh-Triet Tran, Graham Healy, Klaus Schoeffmann, and Cathal Gurrin. 2023. Comparing Interactive Retrieval Approaches at the Lifelog Search Challenge 2021. *IEEE Access* 11 (2023), 30982–30995. doi:10.1109/ACCESS.2023.3248284

[24] Ly-Duyen Tran, Manh-Duy Nguyen, Binh Nguyen, Hyowon Lee, Liting Zhou, and Cathal Gurrin. 2022. E-Myscéal: embedding-based interactive lifelog retrieval system for LSC'22. In *Proceedings of the 5th Annual on Lifelog Search Challenge*. 32–37.

[25] Ly Duyen Tran, Thanh-Binh Nguyen, Cathal Gurrin, and Liting Zhou. 2024. MyEachtraX: Lifelog Question Answering on Mobile. In *Proceedings of the 7th Annual ACM Workshop on the Lifelog Search Challenge* (Phuket, Thailand) *(LSC '24)*. Association for Computing Machinery, New York, NY, USA, 93–98. doi:10.1145/3643489.3661128

[26] Quang-Linh Tran, Binh Nguyen, Gareth Jones, and Cathal Gurrin. 2024. MemoriEase at the NTCIR-17 Lifelog-5 Task. (2024).

[27] Quang-Linh Tran, Binh Nguyen, Gareth J. F. Jones, and Cathal Gurrin. 2024. MemoriEase 2.0: A Conversational Lifelog Retrieve System for LSC'24. In *Proceedings of the 7th Annual ACM Workshop on the Lifelog Search Challenge* (Phuket, Thailand) *(LSC '24)*. Association for Computing Machinery, New York, NY, USA, 12–17. doi:10.1145/3643489.3661114

[28] Gia Huy Vuong, Van-Son Ho, Tien-Thanh Nguyen-Dang, Xuan-Dang Thai, Thang-Long Nguyen-Ho, Minh-Khoi Pham, Tu-Khiem Le, Van-Tu Ninh, and Minh-Triet Tran. 2023. Interactive Sub-Task Focus: LifeInsight's Contribution to NTCIR-17 Lifelog-5. *NII Institutional Repository* (2023). doi:10.20736/0002001333

[29] Rong Yan, Alexander Hauptmann, and Rong Jin. 2003. Multimedia search with pseudo-relevance feedback. In *Image and Video Retrieval: Second International Conference, CIVR 2003 Urbana-Champaign, IL, USA, July 24–25, 2003 Proceedings 2*. Springer, 238–247.

[30] Liting Zhou, Graham Healy, Ly Duyen, Tran, Hideo Joho, Tianbo Ji, and Duc-Tien Dang-Nguyen. [n. d.]. Overview of the NTCIR-17 Lifelog-5 Task. https://api.semanticscholar.org/CorpusID:266493788

[31] Liting Zhou, Minh-Quang Le, Hoang Bao Le, Duc-Tien Dang-Nguyen, Graham Healy, ChenYang Lyu, Tianbo Ji, Hsin-Hung Chen, Longyue Wang, Hideo Joho, Ly-Duyen Tran, Klaus Dajung Jung, Alice Schoeffmannand, Thanh Binh Nguyen, and Cathal Gurrin. 2025. Overview of the NTCIR-18 Lifelog-6 Task. In *Proceedings of the 18th NTCIR Conference on Evaluation of Information Access Technologies (NTCIR-18)*. Tokyo, Japan.

Table 2: Questions and answers for QA task

ID	Question	Ground-truth answer	Predicted answer
QA001	I had a healthy birthday celebration with a lot of fruit and a birthday cake. List the fruits were on the table.	Pineapple, melon, plums, watermelon	Oranges, apples, and bananas.
QA002	What religious landmarks did I see during my trips to Thailand?	Umong Temple	Buddhist temples and stupas
QA003	What is the brand of the drum set in my sister's house?	Millenium	Millenium
QA004	What was the weather like when I last went hiking?	Cloudy	Cloudy
QA005	How many times did I play with my train set in 2019?	17 times	13 times
QA006	How many days did I cycle in April 2020?	22 days	18 days
QA007	When was I last playing the guitar?	June 6, 2020	23 May 2020
QA008	How many days did I have my Christmas tree up at home before I took it down for Christmas 2019?	27 days, from 1/12 to 28/12	23 days from 5/12 to 28/12
QA009	What is the longest duration I ran on the treadmill in one session at a gym?	28 minutes	43 minutes
QA010	When did I last walk along the beach?	24/05/2020	24/05/2020
QA011	In what location do I prefer to run, when I do run?	St. Anne's Park	Outdoors
QA012	I visited a Harry Potter franchise souvenir shop in the UK. What was the name of the shop?	The Boy Wizard or Harry Potter	Don't know
QA013	In my home, there is a staircase. What is the color of this carpet?	Red or wine	Red
QA014	What are the colors of the two long rabbits that live with me at home?	Green and blue	Green and blue
QA015	What brand of TV do I have in the living room at home?	Samsung	Can't determine
QA016	List no more than five brands of beers I have in my fridge, except Guinness.	Bulmers, Heineken, Budweiser, also possible to be; Hope, Moretti, Wicklow Wolf	Heineken, Bulmers
QA017	What date did I go on the longest drive during COVID?	19 April, 2020	14/06/2020
QA018	How many times did I have a picnic by the sea?	2 times, 24, May & 30, June 2020	2 times
QA019	I am overdue a polish and scale. When was the last time I went to see my dentist?	Tuesday, 21 January 2020	Can't determine
QA020	What vitamins do I take regularly?	Multivitamin & vitamin B Complex.	No information
QA021	How many people were in my sister's house on Christmas Day 2019.	8	Can't determine
QA022	Where was I when I was walking with snow on the ground in 2019, outside of Ireland?	Greece or Thesaloniki	Thessaloniki, Greece
QA023	Where did I go on my last business trip in 2019?	Madrid, Spain	Dublin, Ireland
QA024	How many Turkish Airlines flights did I take in 2019?	20	2