MemoriEase 2.0: A Conversational Lifelog Retrieve System for LSC'24

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

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

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

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

ABSTRACT

Lifelog retrieval plays an important role in memory support for lifeloggers. It helps the lifeloggers to browse, search and navigate their life moments from the lifelog data. However, the volume and variety of lifelog data are enormous and range in multiple modalities so they impose a big challenge to retrieve accurate lifelog moments. The Lifelog Search Challenges (LSCs) are a benchmark challenge for evaluating lifelog retrieval systems in different tasks. In this paper, we introduce the MemoriEase 2.0 lifelog retrieval system that participates in LSC'24. This system not only inherits core functions from the precedent system but also incorporates new components such as conversational search, visual similarity search and retrieval-augmented generation for question-answering tasks. The new functions are expected to help expert and novice users solve all topics in three tasks of LSC'24. We evaluate MemoriEase 2.0 in KIS topics in LSC'23 and the system achieves promising results with Recall@1 is 40% at the first hint and it solves 8 over 10 topics.

CCS CONCEPTS

• **Information systems → Users and interactive retrieval**.

KEYWORDS

Lifelog Retrieval, Conversational Search, Personal Archive

ACM Reference Format:
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 *The 7th Annual ACM Lifelog Search Challenge (LSC '24), June 10, 2024, Phuket, Thailand*. ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3643489.3661114

1 INTRODUCTION

Lifelogging is an automatic process of collecting and storing data about an individual daily life, which can be referred to as a lifelogger [8]. Various data across different modalities can be collected such as images/videos from wearable cameras, audio from mobile phones, biometrics from smartwatches, and location from GPS devices. The lifelog data then can be stored in local premises or cloud services to facilitate many applications. Lifelog data can be used to monitor the health of lifeloggers by utilizing biometrics, location and visual data [3, 5]. Another important application of lifelog is a retrieval system for memory support and memory enhancement [15]. Lifelog retrieval systems [22, 29] allow lifeloggers to browse and search for moments in the lifelog dataset, which helps them reminiscence past moments or find some missing information. However, due to the nature of lifelog in collecting over a long period and the variety and repetition of lifelog data, it makes lifelog retrieval challenging to quickly retrieve accurate data.

Lifelog Search Challenges (LSCs) [7] is an annual competition in lifelog retrieval. LSC'24 [9] is the seventh edition of the challenge and the second time that MemoriEase has taken part in it. LSC'24 provides a lifelog dataset and set of topics for participants to find correct lifelog moments based on the topic query. This is a benchmark playground for different lifelog systems to compare and share about their system to improve the knowledge of lifelog retrieval. This challenge has three sub-tasks, namely Known-Item Search (KIS), Ad-Hoc Search, and Question Answering (QA). The KIS provides a query for a specific moment in lifelog data and if the system finds one correct image in the moment, that is counted as correct. In the last LSC'23, MemoriEase [34] found 7 correct answers in 10 topics but only ranked 11th because the scoring mechanism of LSCs considers the time and incorrect submission. The Ad-hoc Search requires as many correct answers as possible for a topic because the Ad-hoc topics are a general moment in life. The QA task asks for a textual answer to a question from the lifelog data. MemoriEase ranked in the top 2 in the QA task in LSC'23 thanks to the fast and accurate retrieved results.

With the development of Generative AI, many applications now provide a chat-based to improve efficiency and user-friendliness such as customer support [4], and search engines [14]. Lifelog retrieval systems can also use language models to provide a conversational search for users. The language models play the role of personal assistants for users to retrieve, browse, and ask for their lifelog information. MemoriEase 2.0 provides a conversational

LSC '24, June 10, 2024, Phuket, Thailand

ACM ISBN 979-8-4007-0550-2/24/06
https://doi.org/10.1145/3643489.3661114

search interface for users to chat with an assistant to find lifelog moments and ask about their lifelog. The conversational search interface provides a more user-friendly interface for users to interact with the system and it also facilitates the question-answering between users and assistants.

Along with the conversational search, we also introduce several new features in the MemoriEase 2.0 system. The visual similarity search empowers the vector search to improve the accuracy of retrieving a general moment for the Ad-hoc task. MemoriEase 2.0 integrates the Retrieval-Augmented Generation [17] (RAG) to retrieve lifelog descriptions and answer the question for QA tasks. These new features are experimented with the LSC'23 topics and prove a promising potential to solve the two challenging tasks in LSC'24.

2 RELATED WORK

Lifelog Search Challenges (LSCs) [7] has been organized for the seventh time so it has already attracted a lot of researchers in constructing lifelog retrieval systems. E-LifeSeeker [20] is an embedding-based retrieval system inherited from LifeSeeker 4.0 [21] that already participated in LSCs for four years. The E-LifeSeeker utilizes multiple embedding models such as CLIP [24] pre-trained on LAION-5B dataset [27], BLIP [19], ALIGN [13] and Coca [35] to embed the lifelog images and queries. It also enhances the user interface to be more intuitive for novice users. Lifelens [12] is built on top of LifeSeeker 4.0 but it presents an innovative UI/UX design for LSC'23. MyEachtra [30] is also a familiar system at LSCs, which is improved from MyScéal [31–33], the LSC'20 to LSC'22 winner. MyEachtra focuses on event-based retrieval and proposes methods to deal with open-ended lifelog question answering. This system achieved top 1 in the QA task of the LSC'23.

The LSC'23 overall winner lifeXplore [16, 26] presents a multi-source search approach, in which embedding-based retrieval results are combined with results from traditional content analysis (for objects, concepts, and recognized text). Both Momento 3.0 [1] and Voxento 4.0 [2] employ the CLIP models [24] as the embedding and enhance the user interface for LSC'24. Instead of using an embedding-based retrieval approach like many other competitors, Rossetto et al [25] uses multi-mode clustering for a graph-based retrieval approach. They organize lifelog data in a multi-modal knowledge graph based on cluster hierarchies. The Best of Both Worlds [28] took part in LSC'23 with a desktop virtual reality hybrid interface. It offers users a result exploration in the virtual reality-based vitrivr-VR [10] and query formulation in the web-based desktop user interface vitrivr-ng [10]. Lifelog Discovery Assistant [11] presents the fourth version of the FIRST system in LSC'23, enhancing its capabilities with generative models for predictive queries and indexing image sequences as events to improve search efficiency. Meanwhile, LifeInsight [23] emphasizes spatial insights and query assistance, leveraging the BLIP embedding model for enhanced accuracy.

MemoriEase [34] participated in LSC'23 for the first time. It utilizes concept-based and embedding-based retrieval approaches to find the results from textual queries. The system uses the BLIP model [19] as the core embedding model and Elasticsearch [1] as the database for storing embeddings, visual concepts and metadata. MemoriEase proposes a free-text search for users with a simple but efficient user interface. There is no need for users to specify any filters because the system offers an automatic filter extraction, which helps to reduce the search time significantly but maintains highly accurate results. In LSC'24, MemoriEase 2.0 continue inheriting the previous strength and improves hugely to adapt to specific tasks. The system integrates an upgraded version of BLIP-2 model [18] and proposes conversational search with RAG for question answering. It also provides the visual similarity search for Ad-hoc tasks.

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

3 MEMORIEASE 2.0

This section introduces the MemoriEase system, from the basic search mechanism to advanced conversational search and other functions. Figure 1 illustrates the overall architecture of the system. The blue rectangle depicts the offline phase of the system, including data processing and indexing. Subsection 3.1 provides detailed information about this phase. When the data is indexed in Elasticsearch, users can perform the search in an online phase. Users chat with the system through a chat interface and the system processes that dialogue to formulate a query. The query is embedded by BLIP-2 and performed a dense vector search in Elasticsearch with embeddings from images. The retrieved results and query are formulated as a prompt to send to OpenAI API to get the response. The retrieved results and responses are then displayed on the interface for users. This is an online phase in the retrieval process, which is illustrated in the green box in the figure and subsection 3.2.

3.1 Data Processing and Indexing

The lifelog dataset for LSC'24 comprises an 18-month lifelog from January 2019 to June 2020, which is similar to LSC'23's dataset. There are a total of 725k lifelog images, along with metadata and visual concepts. We utilized the processed dataset from the previous challenge for this LSC. Specifically, we use edge weight summation to remove blurred images that have a weight lower than a threshold. The event segmentation is performed by grouping all images that have similar vector embedding and in a consecutive time.

We extract the 256-dimension embedding of a main image in each event using the BLIP-2 model. This model is an upgraded version of BLIP, which we used in the previous MemoriEase system. To perform the RAG, we also use InstructBLIP to create the description for each event. The prompt used to create the description is "Act as the first person view, describe detailed information in the image starting with I see.". These data along with the metadata such as location, semantic name, and datetime are indexed to Elasticsearch.

3.2 Conversational Search

Conversational search is a way of retrieving information that users chat with the search system and the system provides searched results along with a textual response. This makes the search process more natural and friendly to users. We employ conversational search in the latest MemoriEase 2.0 system to provide a more

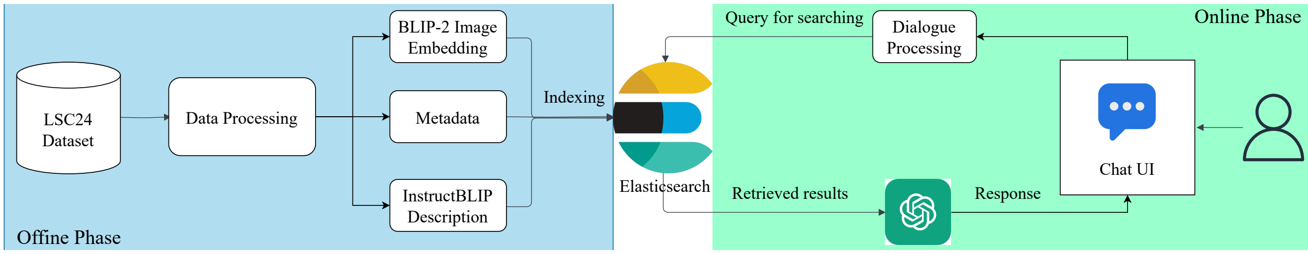

Figure 1: MemoriEase 2.0 Overview

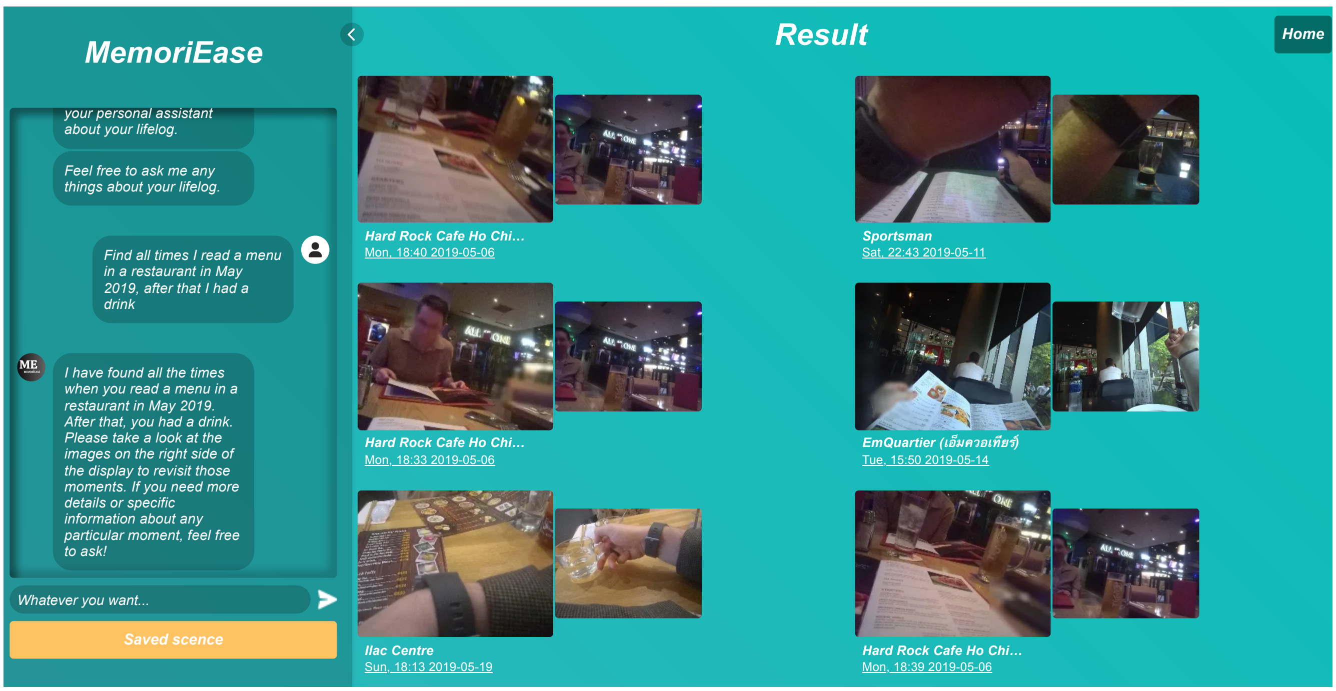

Figure 2: Conversational chat user interface

user-friendly lifelog retrieval system and to address the question-answering task. Users can ask for their lifelog and the system answers the question by retrieving and analyzing a response to the question.

After users submit a chat to the system, the dialogue processing component processes the current and previous chats to aggregate information and produce a single query with metadata filters. The query is embedded into vector embedding and searched in Elasticsearch along with filtering by metadata. The retrieved results and query are formulated to a prompt and sent to a Large Language Model (LLM) to provide a textual response. We use APIs from OpenAI[2] to use the GPT3.5 model as the LLM to analyze the retrieved results and provide the response to users. The system supports automatic temporal search by using a rule-based approach to extract the temporal of the query. The interface for conversational search is depicted in figure 2. After users submit the query "Find all times I read a menu in a restaurant in May 2019, after that I had a drink", the system automatically extracts the filter datetime: "May 2019" and temporal search for the after event: "I had a drink". The retrieved results along with the textual response are displayed in the interface for users to explore.

3.3 Visual Similarity Search

Visual similarity search allows users to search by images and explore lifelog by iteratively refining their input to specific moments. The BLIP-2 embeddings of images are used to calculate the cosine similarity between the input images and the images in the dataset. The top highest relevant images are displayed on the interface for users to click to choose as input and the process is iterated until users find the satisfied results. Users can also submit a query at the beginning to create filters and find the relevant images before retrieving them by visual similarity.

For example, users want to find all images of lifeloggers drinking black beer in 2020. They submit a query and then choose some retrieved images as input and continue the search until they have

[2]https://openai.com/

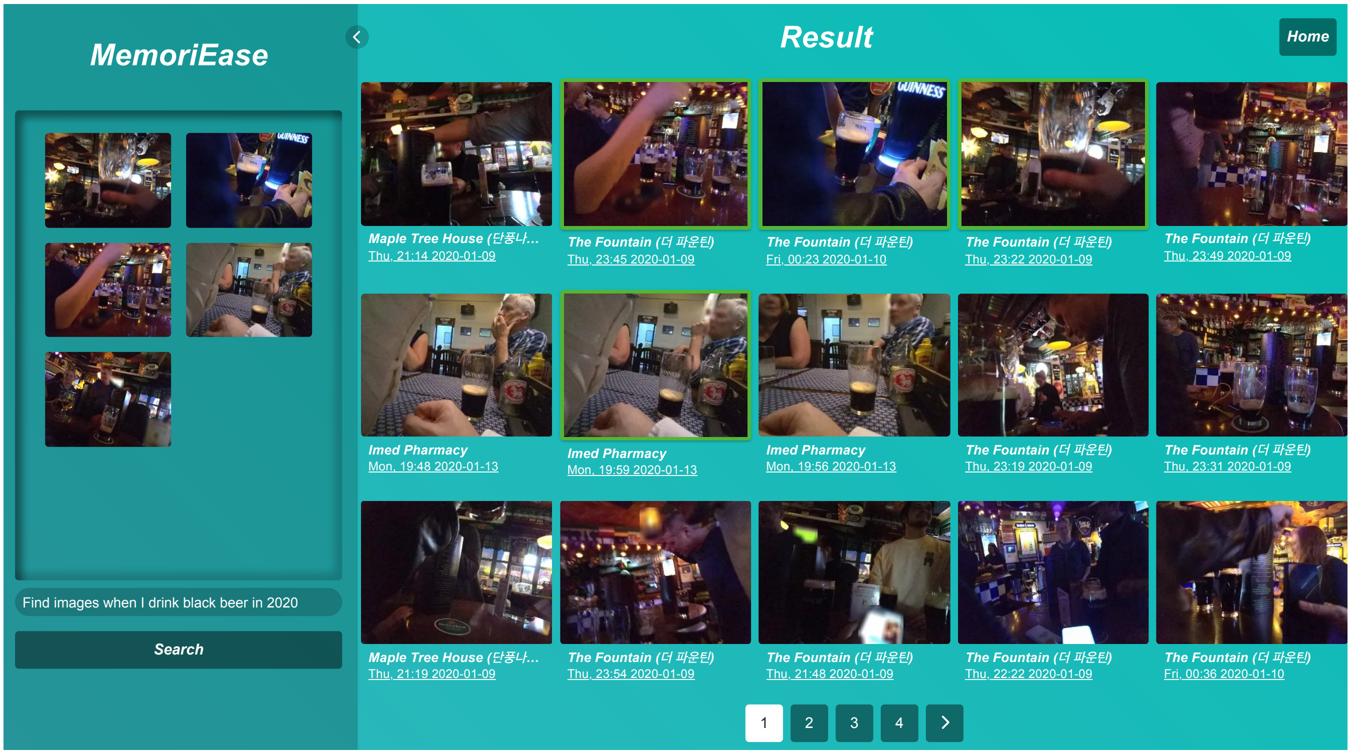

Figure 3: Visual similarity search example

all the images of drinking black beer. Figure 3 illustrates this example. Users choose 5 lifelog images as the input and the system extracts the embedding of these images before aggregating them by averaging. The retrieved results displayed show lifelog images that are similar to the input images.

3.4 RAG for QA tasks

Retrieval-Augmented Generation (RAG) [17] is a technique that helps LLM to incorporate new knowledge from other sources. Lifelog data contains a vast amount of information that can be integrated into LLM by RAG to answer the questions from lifelog. The principle of RAG is simple but efficient. It receives a question as input and then retrieves all the lifelog descriptions that are relevant to the question. The descriptions along with metadata and the question are then formulated to a prompt. The prompt is provided to LLM to get the answer to the question. There are some challenges to getting the correct answer to the question. Firstly, the quality of retrieved results significantly affects the accuracy of LLM's answer. If we provide insufficient or wrong information in the prompt, the answer is incorrect. Secondly, the process of RAG requires several steps, in which retrieving and LLM processing are slow. This makes the system inefficient when time is an important factor in LSCs. Although there are several drawbacks, using RAG for the lifelog QA task is still a promising approach to solving the QA task. We propose a basic pipeline to apply RAG in the MemoriEase 2.0 system. The pipeline is illustrated in figure 4.

InstructBLIP model [6] is a vision-language instruction tuning model based on the pre-trained BLIP-2 models. This shows state-of-the-art performance in multi-modal tasks such as image-text question-answering. We utilize this model to create the description for a main image in each event. The description is indexed to Elasticsearch. We use the GPT3.5 LLM from OpenAI through API requests to find the answer. The answer and retrieved results are displayed to users through the conversational search interface.

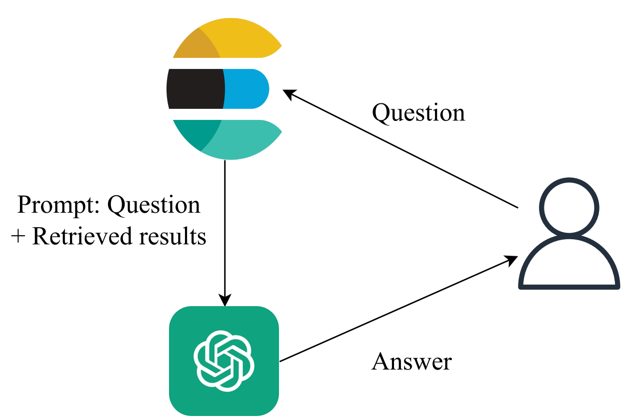

Figure 4: RAG pipeline in MemoriEase 2.0

4 EVALUATION

To evaluate the performance of the BLIP-2 embedding model and conversational chat in the MemoriEase 2.0 system, we use the KIS topics from LSC'23 to measure the Recall@K of different hints in 10 topics. Recall scores at different K from 1 to 100 are calculated to measure the accuracy of retrieved results. There are 6 hints for each topic, and the following hint provides more detailed information than the previous hint. The MemoriEase 2.0 system automatically receives the hints and retrieves results. Table 1 provides the results of the evaluation.

Table 1: Recall at different k in LSC'23 KIS topics

Hints	R@1	R@3	R@5	R@10	R@20	R@50	R@100
1	0.40	0.40	0.40	0.40	0.50	0.60	0.70
2	0.40	0.40	0.50	0.60	0.70	0.70	0.70
3	0.40	0.40	0.50	0.60	0.80	0.80	0.80
4	0.50	0.70	0.80	0.80	0.80	0.80	0.80
5	0.50	0.70	0.80	0.80	0.80	0.80	0.80
6	0.60	0.80	0.80	0.80	0.80	0.80	0.80

As we can see from table 1, the Recall@1 at hint 1 is 40%, which is significantly high. This means that 4 over 10 topics can be found the correct answer at the top 1 results by only in the first hint. However, when more information is provided at hint 6, only 6 over 10 topics are solved. After hint 4, the recall@3 already achieves 70% and increases to 80% at hint 6. When we increase the k, the recall in hint 1 increases up to 70% at k 100. However, in the challenge, there is no chance to scroll up to 100 images in the limited time of 3 minutes for each topic. The maximum recall is 80% at the earliest of hint 6 and k 3.

There are two topics that cannot be found the correct answer, namely LSC23-KIS03 and LSC23-KIS05. The LSC23-KIS03 full hint is "When did I buy that model train? I remember it was a Marklin brand train and I bought it at the weekend. Jer convinced me to buy it when having coffee and I bought it immediately after coffee. It was in June 2019". This topic is in the form of a question and it requires a filter that the MemoriEase 2.0 system not supported is "weekend". It also requires a temporal search in both the previous and following events. This query is complicated and challenges the system to find the correct answer. The LSC23-KIS05 full hint is "Having lunch with Dermot, who was a guest speaker at my lecture. After lunch, he gave a lecture to my class about Lessons in Innovation & Entrepreneurship while I was sitting in the front row. It was in November 2019.". There are several events of having lunch with a man in November 2019 in the dataset, that make the system difficult to retrieve the correct results. In addition, it also requires a temporal search so the system can be failed to perform the temporal search and cannot find the correct answer. These limits should be improved in the future.

5 CONCLUSION

In this paper, we introduce the MemoriEase 2.0 lifelog retrieval system. This system empowers the capability of LLM to construct a conversational search and RAG for the QA task. MemoriEase 2.0 uses the upgraded version BLIP-2 as the core embedding model for the embedding-based retrieval method. Elasticsearch serves as a stable and efficient database for vector and text search. MemoriEase 2.0 provides a new visual similarity search for the Ad-hoc task, that utilizes visual embeddings for better search. These new updates are expected to help users perform well in LSC'24. However, there are still limitations in the system such as RAG speed and efficiency. These drawbacks can be improved in the future to make MemoriEase a user-friendly and powerful lifelog retrieval system.

There are a lot of aspects to improve the MemoriEase lifelog retrieval system in the future. The conversational search requires significant efforts to produce an accurate query aggregated from previous queries. We will improve the dialogue processing component to make the reformulated query more precise in finding the correct lifelog moments. In addition, the visual similarity search introduced in this version still needs more improvements to produce accurate results. This function is an interesting capability of this version that helps explore the lifelog data actively. More filters and embedding models can be used to improve the performance of this function.

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, Yvette Graham, and Cathal Gurrin. 2023. Memento 3.0: An Enhanced Lifelog Search Engine 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, 41–46. https://doi.org/10.1145/3592573.3593103

[2] Ahmed Alateeq, Mark Roantree, and Cathal Gurrin. 2023. Voxento 4.0: A More Flexible Visualisation and Control for Lifelogs. In *Proceedings of the 6th Annual ACM Lifelog Search Challenge* (Thessaloniki, Greece) *(LSC '23)*. Association for Computing Machinery, New York, NY, USA, 7–12. https://doi.org/10.1145/3592573.3593097

[3] Adrienne H Andrew, Kevin Eustice, and Andy Hickl. 2013. Using location lifelogs to make meaning of food and physical activity behaviors. In *2013 7th International Conference on Pervasive Computing Technologies for Healthcare and Workshops*. IEEE, 408–411.

[4] Erik Brynjolfsson, Danielle Li, and Lindsey R Raymond. 2023. *Generative AI at work*. Technical Report. National Bureau of Economic Research.

[5] Andrea LS Bulungu, Luigi Palla, Jan Priebe, Lora Forsythe, Pamela Katic, Gwen Varley, Bernice D Galinda, Nakimuli Sarah, Joweria Nambooze, Kate Wellard, et al. 2021. Validation of a life-logging wearable camera method and the 24-h diet recall method for assessing maternal and child dietary diversity. *British Journal of Nutrition* 125, 11 (2021), 1299–1309.

[6] Wenliang Dai, Junnan Li, Dongxu Li, Anthony Meng Huat Tiong, Junqi Zhao, Weisheng Wang, Boyang Li, Pascale N Fung, and Steven Hoi. 2024. Instructblip: Towards general-purpose vision-language models with instruction tuning. *Advances in Neural Information Processing Systems* 36 (2024).

[7] Cathal Gurrin, Björn Þór Jónsson, Duc Tien Dang Nguyen, Graham Healy, Jakub Lokoc, Liting Zhou, Luca Rossetto, Minh-Triet Tran, Wolfgang Hürst, Werner Bailer, and Klaus Schoeffmann. 2023. Introduction to the Sixth Annual Lifelog Search Challenge, LSC'23. In *Proceedings of the 2023 ACM International Conference on Multimedia Retrieval* (Thessaloniki, Greece) *(ICMR '23)*. Association for Computing Machinery, New York, NY, USA, 678–679. https://doi.org/10.1145/3591106.3592304

[8] 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.

[9] Cathal Gurrin, Liting Zhou, Graham Healy, Bailer. Werner, Duc-Tien Dang-Nguyen, Steve Hodges, Björn Þór Jónsson, Jakub Lokoč, Luca Rossetto, Minh-Triet

Tran, and Klaus Schöffmann. 2024. Introduction to the Seventh Annual Lifelog Search Challenge, LSC'24. *International Conference on Multimedia Retrieval (ICMR'24)*. https://doi.org/10.1145/3652583.3658891

[10] Silvan Heller, Luca Rossetto, Loris Sauter, and Heiko Schuldt. 2022. vitrivr at the Lifelog Search Challenge 2022. In *Proceedings of the 5th Annual on Lifelog Search Challenge* (Newark, NJ, USA) *(LSC '22)*. Association for Computing Machinery, New York, NY, USA, 27–31. https://doi.org/10.1145/3512729.3533003

[11] Nhat Hoang-Xuan, Thang-Long Nguyen-Ho, Cathal Gurrin, and Minh-Triet Tran. 2023. Lifelog Discovery Assistant: Suggesting Prompts and Indexing Event Sequences for FIRST at LSC 2023. In *Proceedings of the 6th Annual ACM Lifelog Search Challenge* (Thessaloniki, Greece) *(LSC '23)*. Association for Computing Machinery, New York, NY, USA, 47–52. https://doi.org/10.1145/3592573.3593104

[12] Maria Tysse Hordvik, Julie Sophie Teilstad Østby, Manoj Kesavulu, Thao-Nhu Nguyen, Tu-Khiem Le, and Duc-Tien Dang-Nguyen. 2023. LifeLens: Transforming Lifelog Search with Innovative UX/UI Design. In *Proceedings of the 6th Annual ACM Lifelog Search Challenge* (Thessaloniki, Greece) *(LSC '23)*. Association for Computing Machinery, New York, NY, USA, 1–6. https://doi.org/10.1145/3592573.3593096

[13] Chao Jia, Yinfei Yang, Ye Xia, Yi-Ting Chen, Zarana Parekh, Hieu Pham, Quoc Le, Yun-Hsuan Sung, Zhen Li, and Tom Duerig. 2021. Scaling up visual and vision-language representation learning with noisy text supervision. In *International conference on machine learning*. PMLR, 4904–4916.

[14] Dominique Kelly, Yimin Chen, Sarah E Cornwell, Nicole S Delellis, Alex Mayhew, Sodiq Onaolapo, and Victoria L Rubin. 2023. Bing Chat: The Future of Search Engines? *Proceedings of the Association for Information Science and Technology* 60, 1 (2023), 1007–1009.

[15] Basel Kikhia, Josef Hallberg, Johan E Bengtsson, Stefan Savenstedt, and Kare Synnes. 2010. Building digital life stories for memory support. *International journal of Computers in Healthcare* 1, 2 (2010), 161–176.

[16] Andreas Leibetseder and Klaus Schoeffmann. 2021. lifexplore at the lifelog search challenge 2021. In *Proceedings of the 4th Annual on Lifelog Search Challenge*. 23–28.

[17] Patrick Lewis, Ethan Perez, Aleksandra Piktus, Fabio Petroni, Vladimir Karpukhin, Naman Goyal, Heinrich Küttler, Mike Lewis, Wen-tau Yih, Tim Rocktäschel, et al. 2020. Retrieval-augmented generation for knowledge-intensive nlp tasks. *Advances in Neural Information Processing Systems* 33 (2020), 9459–9474.

[18] 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.

[19] 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.

[20] Thao-Nhu Nguyen, Tu-Khiem Le, Van-Tu Ninh, Cathal Gurrin, Minh-Triet Tran, Thanh Binh Nguyen, Graham Healy, Annalina Caputo, and Sinead Smyth. 2023. E-LifeSeeker: An Interactive Lifelog Search Engine 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, 13–17. https://doi.org/10.1145/3592573.3593098

[21] 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

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

[23] Tien-Thanh Nguyen-Dang, Xuan-Dang Thai, Gia-Huy Vuong, Van-Son Ho, Minh-Triet Tran, Van-Tu Ninh, Minh-Khoi Pham, Tu-Khiem Le, and Graham Healy. 2023. LifeInsight: An Interactive Lifelog Retrieval System with Comprehensive Spatial Insights and Query Assistance. In *Proceedings of the 6th Annual ACM Lifelog Search Challenge* (Thessaloniki, Greece) *(LSC '23)*. Association for Computing Machinery, New York, NY, USA, 59–64. https://doi.org/10.1145/3592573.3593106

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

[25] Luca Rossetto, Oana Inel, Svenja Lange, Florian Ruosch, Ruijie Wang, and Abraham Bernstein. 2023. Multi-Mode Clustering for Graph-Based Lifelog Retrieval. In *Proceedings of the 6th Annual ACM Lifelog Search Challenge* (Thessaloniki, Greece) *(LSC '23)*. Association for Computing Machinery, New York, NY, USA, 36–40. https://doi.org/10.1145/3592573.3593102

[26] Klaus Schoeffmann. 2023. lifeXplore at the Lifelog Search Challenge 2023. In *Proceedings of the 6th Annual ACM Lifelog Search Challenge* (Thessaloniki, Greece) *(LSC '23)*. Association for Computing Machinery, New York, NY, USA, 53–58. https://doi.org/10.1145/3592573.3593105

[27] Christoph Schuhmann, Romain Beaumont, Richard Vencu, Cade Gordon, Ross Wightman, Mehdi Cherti, Theo Coombes, Aarush Katta, Clayton Mullis, Mitchell Wortsman, et al. 2022. Laion-5b: An open large-scale dataset for training next generation image-text models. *Advances in Neural Information Processing Systems* 35 (2022), 25278–25294.

[28] Florian Spiess, Ralph Gasser, Heiko Schuldt, and Luca Rossetto. 2023. The Best of Both Worlds: Lifelog Retrieval with a Desktop-Virtual Reality Hybrid System. In *Proceedings of the 6th Annual ACM Lifelog Search Challenge* (Thessaloniki, Greece) *(LSC '23)*. Association for Computing Machinery, New York, NY, USA, 65–68. https://doi.org/10.1145/3592573.3593107

[29] 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*. 24–29.

[30] 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

[31] Ly-Duyen Tran, Manh-Duy Nguyen, Nguyen Thanh Binh, Hyowon Lee, and Cathal Gurrin. 2020. Myscéal: an experimental interactive lifelog retrieval system for LSC'20. In *Proceedings of the Third Annual Workshop on Lifelog Search Challenge*. 23–28.

[32] 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* (Newark, NJ, USA) *(LSC '22)*. Association for Computing Machinery, New York, NY, USA, 32–37. https://doi.org/10.1145/3512729.3533012

[33] 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.

[34] 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*. 30–35.

[35] Jiahui Yu, Zirui Wang, Vijay Vasudevan, Legg Yeung, Mojtaba Seyedhosseini, and Yonghui Wu. 2022. Coca: Contrastive captioners are image-text foundation models. *arXiv preprint arXiv:2205.01917* (2022).