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/10.20736/0002001333

Interactive Sub-Task Focus: LifeInsight's Contribution to NTCIR-17 Lifelog-5

Gia-Huy Vuong
University of Science, VNU-HCM
Vietnam National University
Ho Chi Minh City
Vietnam

Van-Son Ho
University of Science, VNU-HCM
Vietnam National University
Ho Chi Minh City
Vietnam

Tien-Thanh Nguyen-Dang
University of Science, VNU-HCM
Vietnam National University
Ho Chi Minh City
Vietnam

Xuan-Dang Thai
University of Science, VNU-HCM
Vietnam National University
Ho Chi Minh City
Vietnam

Thang-Long Nguyen-Ho
University of Science, VNU-HCM
Vietnam National University
Ho Chi Minh City
Vietnam

Minh-Khoi Pham
Dublin City University
Ireland

Tu-Khiem Le
Dublin City University
Ireland

Van-Tu Ninh
Dublin City University
Ireland

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

ABSTRACT

The rise of digital storage technology and portable sensors has led to an increase in lifelogging, where individuals digitally record their personal experiences. This has opened up new research opportunities in lifelog data retrieval. However, the real-time and automatic recording of data by sensors presents unique challenges compared to traditional search engines, particularly in data organization and search. The highly personalized nature of the dataset also necessitates the consideration of user interactions and feedback in the search engine. In this paper, we present LifeInsight, a robust lifelog retrieval system designed specifically for the NTCIR17 Lifelog-5 Task. Originally developed for the Lifelog Search Challenge (LSC), the system has been adapted and optimized to address the unique requirements of the Lifelog Semantic Access Task (LSAT). Of the two tasks within NTCIR17 Lifelog-5, our primary focus is on the interactive sub-task, which involves evaluating LifeInsight's performance under different user interaction approaches employed by various users. Therefore, a comprehensive user study was conducted to evaluate the LifeInsight system encompassed both expert and novice users across various settings, including ad-hoc and known-item-search scenarios.

KEYWORDS

lifelog, information retrieval, quantified self, personal data

TEAM NAME

HCMUS

SUBTASKS

Lifelog Semantic Access Task (LSAT)

1 INTRODUCTION

Lifelog moment retrieval is a specialized branch of information retrieval that concentrates on lifelog data. These data are comprehensive archives of personal experiences, captured through various sensors in everyday life, and encompass a multitude of modalities. Numerous lifelog moment retrieval challenges have been established to stimulate the research community to propose innovative solutions for this burgeoning field of study [7, 9, 12, 17]. Lifelog moment retrieval, the process of identifying relevant lifelog moments in response to a specific query, presents a unique set of challenges. Lifelog spans different types of data from egocentric images to precise location information, activity, semantic physiological data. Despite these complexities, the field is experiencing rapid growth and evolution, with numerous strategies being explored for effective lifelog retrieval. Among the most promising directions of research are the application of deep learning techniques and the utilization of multimodal features. Additionally, lifelog moment retrieval challenges require participants to find images in their own lifelog data that are relevant to a specific information need.

The fifth NTCIR Lifelog task was successfully conducted at NTCIR-17 [31], continuing the tradition established in NTCIR-12, NTCIR-13, NTCIR-14, and NTCIR-16 [8–10, 32]. Like previous NTCIR Lifelog Tasks [8–10, 32], the NTCIR-17 Lifelog-5 Task [31] focuses on the Lifelog Semantic Access Task (LSAT). This is a known-item search task where participants must retrieve specific moments in a lifelogger's life. The task can be approached either automatically or interactively. In simpler terms, the task is to find specific images in a lifelog collection that match specific queries. For example, a query might be "Find all images of me at my wedding" or "Find all images of my cat playing with a ball". Nevertheless, the NTCIR-17 Lifelog-5 task is different from the Lifelog Search Challenge (LSC) [11] in several ways, even though both competitions allow interactive systems to solve queries. The latter considers search time as a fundamental aspect of its scoring mechanism, while the NTCIR-17 Lifelog-5 Task focuses more on the accuracy of the search engine rather than its retrieval time. Moreover, the NTCIR-17 Lifelog-5 Task requires participants to retrieve all relevant images, as opposed to the LSC [11] only requires participants to retrieve a single correct image for a given query. In other words, the LSC is a more focused retrieval task, while the NTCIR-17 Lifelog-5 task is a more comprehensive retrieval task. This difference in requirements reflects the different goals of the two challenges. In addition, the Lifelog Semantic Access Task (LSAT) in NTCIR-17 is the one that has the largest query bank for the lifelog moment retrieval task among multiple different lifelog search challenges, with 41 queries in total. It consists of two subtasks: an ad-hoc subtask with 21 queries, and a known-item search subtask with 24 queries. In the ad-hoc subtask, users try to find as many relevant images as possible for a given topic while in the known-item search subtask, they need to find one or a few images that match a given description precisely.

In summary, we adapted the comprehensive lifelog retrieval system LifeInsight which was originally developed for the Sixth Annual ACM Lifelog Search Challenge [11], to the NTCIR-17 Lifelog-5 LSAT subtask by adding the following enhancement features:

(1) **Automatic query parser**: LifeInsight can now automatically extract concept information from a query, such as location, object, and negative prompt. This benefits Elastic by helping it to find all relevant images more effectively.
(2) **Visual example generation**: LifeInsight enables users to search with generated images from Stable Diffusion [24]. This feature helps the system find relevant results more easily through visual search.

2 RELATED WORK

The lifelog retrieval field has made remarkable strides in recent years, culminating in the creation of competitions to encourage the development of interactive retrieval systems. These systems aim to swiftly pinpoint specific images from a vast lifelog data collection within a set time frame for a query. The Lifelog Search Challenge [11] is a prominent competition in this area, garnering significant interest and participation.

Various systems, including Memoria[memoria], LifeSeeker [20], vitrivr [14], FIRST [30], and Myscéal [28], have offered multiple search modalities based on concepts. Lifegraph [25] and LifeConcept [4] utilized knowledge graphs and concept recommendation methods like ConceptNet to facilitate retrieval by linking relevant concepts with images. Other systems such as lifeXplore [16], PhotoCube [27], and LifeMon [6] employed convolutional neural networks (CNNs) like YOLOv4 [5] and traditional object detectors for content analysis. These systems primarily used Database Management Systems (DBMS) or Elasticsearch data retrieval mechanisms to effectively align user queries with visual concepts and metadata. A number of systems, including LifeSeeker 4.0 [19], E-Myscéal [29], Memento 2.0 [2], FIRST 3.0 [15], and Voxento [3], incorporated vision-language pre-trained models, specifically the CLIP model [22]. These systems demonstrated significant performance improvements in zero-shot image-text retrieval compared to their previous versions.

In the previous NTCIR-16 event [32], three teams proposed their interactive retrieval systems for the Lifelog Semantic Access Task (LSAT). Firstly, the DCU and HCMUS [18] team developed two lifelog data retrieval systems for the NTCIR-16 Lifelog-4 Task [32], which are LifeSeeker and Myscéal. Both systems were evaluated in both automatic and interactive settings, with Myscéal surpassing LifeSeeker in the interactive setting. The study concluded that lifelog retrieval has the potential to be a powerful tool for personal data management and analysis. Secondly, two interactive retrieval systems, DCUMemento and DCUVOX [1], were proposed for the NTCIR-16 Lifelog-4 Task [32]. Both systems use image-text embeddings to build their search backend from various CLIP [23] models. The paper also discusses the query reformulation strategy used by the systems and presents the results of their evaluation. The systems have some limitations, but improvements are planned for future iterations. Additionally, the THUIR-LL team [13] developed an enhanced interactive lifelog search engine for the NTCIR-16 Lifelog-4 Task [32]. The search engine includes a query text parsing procedure, a feedback mechanism with ternary feedback and negative keywords, and a result presentation for interaction that shows relevant images in a T-shape fixation distribution with timeline viewing. The experiment on constructed topics shows promising progress for both novice and expert users, and the online evaluation results indicate the usefulness and precision of the search engine.

3 LIFEINSIGHT AT THE NTCIR-17 LIFELOG-5

LifeInsight[21], a system originally developed for the Sixth Annual ACM Lifelog Search Challenge [11], was adapted to deal with the interactive subtask of the NTCIR-17 Lifelog-5 task. This adaptation involved adding two features: an automatic query parser and a visual example generator, both of which are described in 3.2. 3.1 provides an overview of LifeInsight.

3.1 An overview of LifeInsight system

LifeInsight [21] is an advanced interactive lifelog retrieval system with context awareness. It aims to provide users with a comprehensive and efficient way to retrieve and analyze lifelog data, offering to search for images based on visual input or text. A brief overview of LifeInsightcan be illustrated in Fig 1.

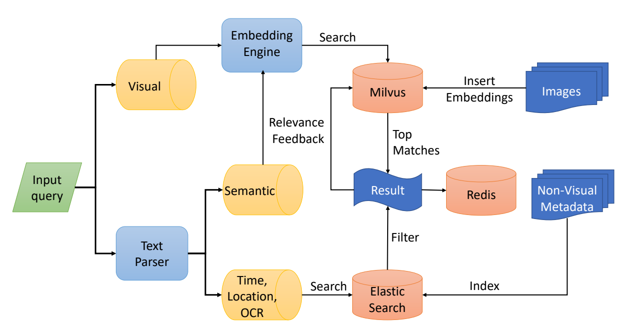

Figure 1: The overview architecture of the LifeInsight system.

Our retrieval system stands out with its unique and user-friendly interface, setting it apart from conventional systems. Instead of the traditional search bar and image result gallery, our interface presents an image gallery that displays search results in a variety of cluster view modes. The system also includes a multi-functional chatbox that not only maintains a history of queries but also supports both text and image inputs. For more advanced actions, users can employ commands. The interface also features a vertical navigation panel equipped with handy buttons for easy access to key features. These include the ability to hide or reveal the chatbox, view the spatial insight mAP, provide feedback, and switch between gallery view modes. Figure 2 illustrates the main user interface of our retrieval system.

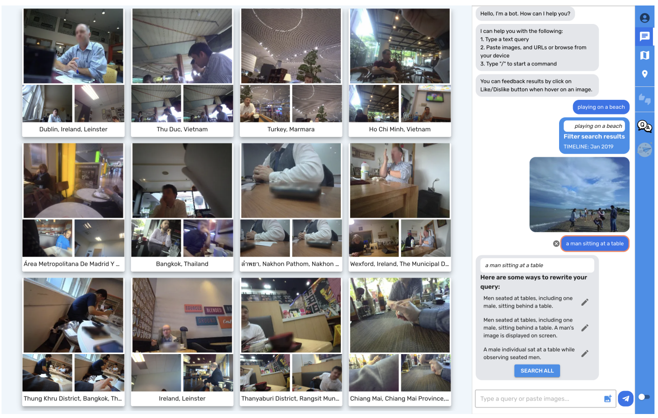

Figure 2: The user interface of the LifeInsight system.

3.2 LifeInsight at the NTCIR-17 Lifelog-5

As previously discussed, we have developed an Automatic Query Parser to address the interactive subtask of the NTCIR-17 Lifelog-5. This parser is capable of automatically extracting concept information from a query, including elements such as location, object, transportation, color, or negative prompts. The inclusion of this concept information significantly enhances the performance of our system.

In addition, our system introduces the innovative feature of Visual Example Generation, designed to assist users in visualizing their desired images. Utilizing a cutting-edge text-to-image model, such as the Stable Diffusion model [24], the system analyzes user queries and generates images that best align with the user's search criteria. This feature bridges the gap between the user's mental image and the actual search query, facilitating a more accurate envisioning of the target images. Our system's visual search capability allows users to select an image or a group of images that match their interests and ask the system to retrieve similar images. This feature proves especially beneficial for those seeking images that are challenging to articulate in words. To illustrate the practicality of the Visual Example Generating feature, let's take a scenario into account. Imagine we're trying to locate a specific image displayed on the left side of Figure 3. We could use the following description to generate example images: "A first-person view in a Greek restaurant, drinking a small bottle of wine, and eating Greek food (chips and meat) with a salad". Although users could potentially locate the target image through free-text search, it might necessitate scrolling through numerous search results. However, with the Visual Example Generating feature, our system can produce four images based on the description in approximately 10 seconds. This allows users to carry on with other searches while waiting for the generated examples.

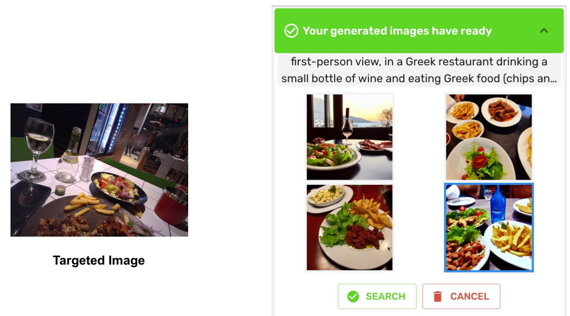

Figure 3: An example of the Visual Example Generation function in the LifeInsight system.

4 EXPERIMENT

To begin with, the HCMUS team participated in the NTCIR 17 Lifelog-5 Interactive sub-task and developed the lifelog data retrieval system LifeInsight. We conducted an interactive performance evaluation of LifeInsight involving 8 participants. The study was designed to assess the system's effectiveness in retrieving lifelog data in both ad-hoc and known-item topic settings. Among the participants, there was one expert user who had previously participated in the Lifelog Challenge and had experience with lifelog retrieval problems. The remaining seven were novice users, who were encountering lifelog retrieval problems for the first time.

Training an expert user on all the features of LifeInsight took approximately five minutes, enabling them to effectively handle queries from the challenge. For novice users, the training period did not exceed fifteen minutes. This included an explanation of lifelog retrieval problems and their benefits, which helped to spark their interest in the challenge and potentially enhance their performance in LifeInsight.

Furthermore, to enhance the reliability of LifeInsight's performance, we introduced two vision-language pre-trained models, BLIP-2 and CLIP, both with 768 dimensions. We randomly assigned each user one of the two models based on their name, ensuring that they were unaware of the specific model being used. This allowed us to perform benchmarking upon receiving their submissions.

4.1 Experimental Setup

We performed two interactive search runs, one for an expert user and one for novice users. For each query, the user had a few minutes to read the query and any relevant information before the countdown clock started from 300 seconds. As soon as the user found a relevant image on the result page, they could submit it using the submit button. If the user submitted a thumbnail of a scene (i.e., a series of similar images) in the temporal browsing view, all of the images in the scene were submitted at the same time. LifeInsight enables users to swiftly submit up to 100 images for a single query. To prevent spamming the challenge server, the system incorporates a random sleep feature during the submission process. For a more professional approach, we have deployed these queries, which include two types: ad-hoc search and known-item search, on the DRES system [26].

In our experiments, we evaluated the performance of our information retrieval system, LifeInsight, across several runs, each identified by a unique code (e.g., U1-A1, U2-A1, etc.). The code U1-A1 indicates that the run was performed by user ID 1 using approach 1. For each run, we recorded the following metrics:

- **Number of queries attempted:** This is the total number of queries that the user submitted.
- **Number of images correctly submitted:** The number of relevant images that were retrieved.

In addition to these metrics, we also calculated several standard information retrieval metrics, including:

- **Mean Average Precision (mAP)**: This metric provides an aggregate measure of retrieval performance, computed as the mean of precision values obtained after each relevant document is retrieved. It offers a comprehensive evaluation of the ranking quality of the retrieval system.
- **Precision at 5 (P@5)**: This metric evaluates the effectiveness of the retrieval system in returning highly relevant results within the top 5 images. It quantifies the proportion of relevant images among the top 5 retrieved, thereby assessing the precision of the system in its most immediate results.
- **Mean Precision at 10 (P@10)**: Similar to P@5, this metric extends the evaluation to the top 10 images retrieved by the system. It measures the proportion of relevant images within these top 10 results, providing insight into the system's precision in a broader retrieval context.

These metrics are particularly useful when we are interested in the topmost results returned by our search system, which is often the case in Information Retrieval. A higher value for these metrics indicates that more relevant images are being returned at the top of our search results.

4.2 Experimental Result

The table 1 shows the evaluation of the interactive LifeInsightsystem on 8 users, including 1 expert user and 7 novice users, using two approaches: the BLIP model and the CLIP model.

4.2.1 Approach 1 (BLIP model):

- **Expert**: The expert user attempted 40 queries and retrieved 932 images, 350 of which were relevant. His mAP was 0.164, indicating a moderate level of precision across all queries. His P@5 and P@10 were 0.32 and 0.245 respectively, suggesting that the most relevant images were ranked higher in his search results.
- **U1-A1**: This user attempted the same number of queries as the expert but had a lower performance. He retrieved more images (951), but fewer of them were relevant (213). His mAP was 0.0614, and his P@5 and P@10 were 0.215 and 0.1725 respectively.
- **U2-A1**: This user attempted fewer queries (37) but retrieved the most images (1221), of which only 226 were relevant. His mAP was 0.1687, and his P@5 and P@10 were 0.2162 and 0.1676 respectively.
- **U3-A1**: This user attempted fewer queries (36) but had a slightly better performance than U2-A1. He retrieved fewer images (1181), but more of them were relevant (374). His mAP was 0.127, and his P@5 and P@10 were 0.3056 and 0.2528 respectively.
- **U4-A1**: This user attempted more queries (38) and had a performance that was between U2-A1's and U3-A1's. He retrieved 1024 images, 282 of which were relevant. His mAP was 0.1255, which was higher than all users except the expert. His P@5 and P@10 were also higher than all users except the expert.

While all users attempted a substantial number of queries and retrieved a significant number of images, there was a noticeable variation in their ability to retrieve relevant images and rank them effectively. The expert user, despite attempting the same number of queries as U1-A1, demonstrated superior performance across all metrics, retrieving 932 images with 350 being relevant. This is reflected in their mAP score of 0.164 and P@5 and P@10 scores of 0.32 and 0.245 respectively. This highlights the expert's superior information retrieval skills compared to the other users (as shown in Fig 4).

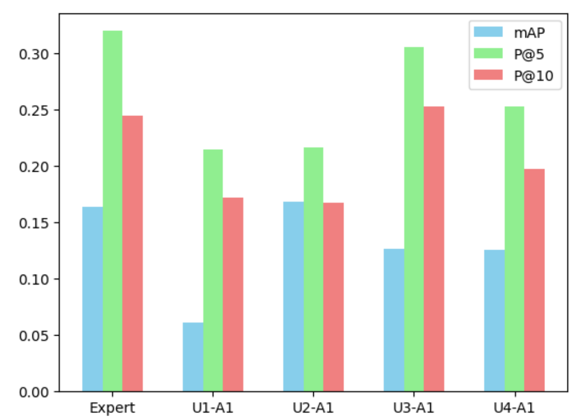

Figure 4: Evaluation of the interactive LifeInsight system on five users with Approach 1 – BLIP model.

4.2.2 Approach 2 (CLIP model):

- **U1-A2**: This user attempted 35 queries and retrieved 1285 images, 291 of which were relevant. His mAP was 0.129, indicating a moderate level of precision across all queries. His P@5 and P@10 were 0.2514 and 0.2286 respectively.

Run	# queries attempted	# images correct/submitted	mAP	P@5	P@10
Expert	40	350/932	0.164	**0.32**	**0.245**
U1-A1	40	213/951	0.0614	0.215	0.1725
U2-A1	37	226/1221	**0.1687**	0.2162	0.1676
U3-A1	36	374/1181	0.127	0.3056	**0.2528**
U4-A1	38	282/1024	0.1255	0.2526	0.1974
U1-A2	35	291/1285	**0.129**	0.2514	0.2286
U2-A2	36	182/1121	0.0872	0.2056	0.1611
U3-A2	39	208/870	0.1276	**0.2923**	**0.2308**

Table 1: Results of LifeInsight's iterative runs with users.

- **U2-A2**: This user attempted more queries (36) but retrieved fewer images (1121), of which even fewer were relevant (182). His mAP was 0.0872, which is lower than U1-A2's. His P@5 and P@10 were also lower than U1-A2's.
- **U3-A2**: This user attempted the most queries (39) but retrieved the least number of images (870). However, more of them were relevant (208) compared to U2-A2. His mAP was 0.1276, which was slightly lower than U1-A2's but higher than U2-A2's. His P@5 and P@10 were also higher than both U1-A2's and U2-A2's.

In short, while U1-A2 retrieved the most images, his ability to retrieve relevant images was not as good as U3-A2's. On the other hand, even though U3-A2 retrieved fewer images, he had a higher mAP, P@5, and P@10 compared to U2-A2, indicating that he was more effective at retrieving relevant images and ranking them higher in his search results (as shown in Fig 5).

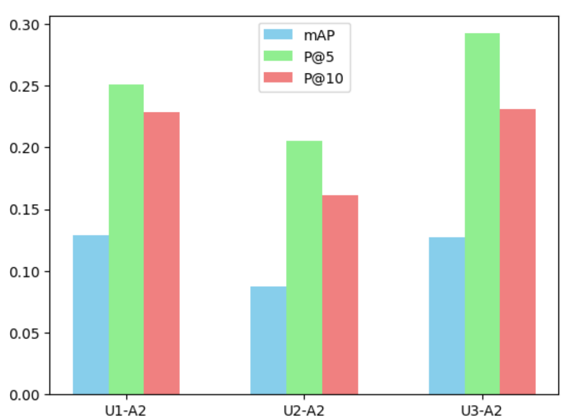

Figure 5: Evaluation of the interactive LifeInsight system on three users with Approach 2 – CLIP model.

4.2.3 Comparison between Approaches: In the A1 approach, the expert and users U1-A1, U2-A1, U3-A1, and U4-A1 participated. The expert exhibited superior performance across all metrics. The number of images retrieved in this approach varied greatly, with the expert retrieving 932 images and U2-A1 retrieving the most at 1221. The relevance of the retrieved images also varied, with the expert retrieving the highest number of relevant images (350). In terms of precision, the expert achieved the highest Mean Average Precision (mAP) and mean precision at ranks 5 and 10.

Conversely, in the A2 approach, users U1-A2, U2-A2, and U3-A2 were involved. Among this group, U3-A2 outperformed the others. This approach resulted in fewer retrieved images compared to Approach A1, with numbers ranging from 870 by U3-A2 to 1285 by U1-A2. The number of relevant retrieved images was also less than in Approach A1, with U1-A2 retrieving the most relevant images (291) in this group. However, in terms of precision metrics such as mAP and mean precision at ranks 5 and 10, U3-A2 had the highest scores among all users in this group.

In conclusion, while Approach A1 led to a higher number of retrieved and relevant images, it required expertise to achieve high precision. Conversely, Approach A2 resulted in fewer retrieved images but had comparable precision metrics even without an expert user. Therefore, Approach A1 might be more suitable for experts who can effectively sift through more information, while Approach A2 might be more user-friendly for non-experts. Further research could explore ways to combine the strengths of both approaches.

4.2.4 Conclusion: As depicted in Table 1, the system effectively assisted novice users in addressing a substantial number of queries, even without prior experience. The expert user set a high benchmark by attempting 40 queries and accurately identifying 350 out of 932 retrieved images. In contrast, novice users, on average, attempted 37 queries and correctly identified a varying number of images.

While other metrics, such as Mean Average Precision (mAP) and the number of attempted queries, are also important, the combination of three key factors underscores U3-A1 as the most proficient novice user in generating pertinent results. In essence, U3-A1 stands out as the top-performing novice user. Although the expert user surpassed the best novice user in terms of precision and accuracy in retrieving relevant images, it is noteworthy that the top novice user achieved a mAP of 0.1687. This indicates that our system is successful in bridging the gap between expert and novice users, demonstrating its effectiveness even when users lack prior experience (as shown clearly in Figure 6).

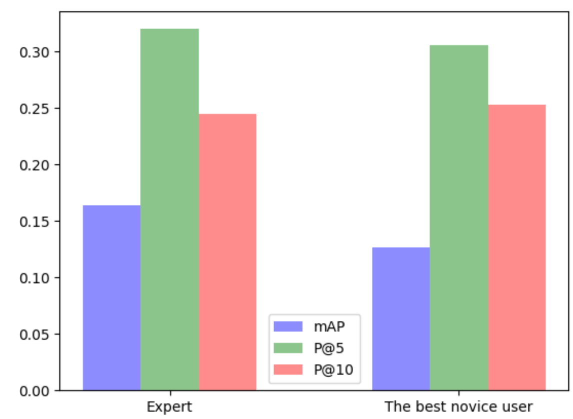

Figure 6: Comparative performance analysis between the best novice user and the expert user.

4.3 Users Experience Questionnaire

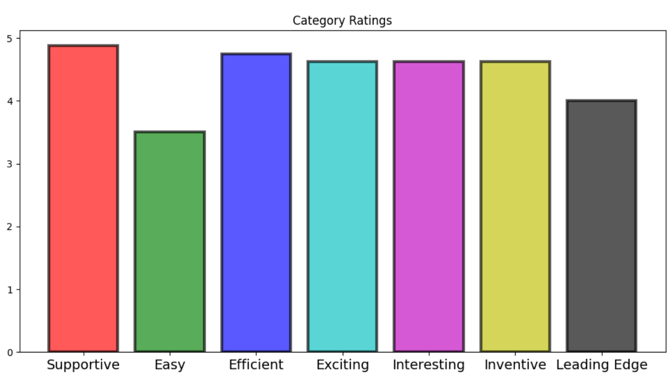

Figure 7: Average user experience ratings across different categories.

In our research, we evaluated seven different aspects of our system: 'Supportive', 'Easy', 'Efficient', 'Exciting', 'Interesting', 'Inventive', and 'Leading Edge'. The system was rated on a scale of 1 to 5 for each aspect by multiple users. The average scores were calculated (shown in Fig 7) and are as follows:

The system was found to be highly 'Supportive' with an average score of 4.88, indicating that users felt well-assisted throughout their interaction. The 'Easy' aspect had an average score of 3.5, suggesting that while some users found the system easy to use, others faced challenges. The system was rated highly 'Efficient' with an average score of 4.75, demonstrating its ability to deliver results promptly and accurately. In terms of being 'Exciting', the system scored an average of 4.63, showing that users found their interaction with the system engaging and stimulating. The system was considered 'Interesting' with an average score of 4.63, reflecting that users found the system intriguing and were keen to explore more. The 'Inventive' aspect had an average score of 4.63, indicating that users appreciated the innovative approach of the system in solving tasks. Lastly, the system was rated 4.13 on being 'Leading Edge', suggesting that while many users considered the system to be at the forefront of technology, there is still room for improvement. These results provide valuable insights into how our system is perceived by users and will guide future improvements to enhance user experience.

5 CONCLUSION

In conclusion, this study demonstrated the efficacy of a system in aiding novice users to address a significant number of queries, even with minimal training. The system achieved a high benchmark, with expert users excelling in query resolution and image identification. Novice users also showed promising results, successfully handling an impressive percentage of queries after less than 15 minutes of training. Moreover, the system's effectiveness suggests its potential to enhance productivity among novice users. Future research could focus on designing more personalized systems that offer detailed feedback and guidance, thereby further improving the assistance provided to novice users.

REFERENCES

[1] Naushad Alam and Ahmed Alateeq. [n. d.]. DCU at the NTCIR16 Lifelog-4 Task. https://api.semanticscholar.org/CorpusID:251474958
[2] Naushad Alam, Yvette Graham, and Cathal Gurrin. 2022. Memento 2.0: An Improved 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, 2–7. https://doi.org/10.1145/3512729.3533006
[3] Ahmed Alateeq, Mark Roantree, and Cathal Gurrin. 2022. Voxento 3.0: A Prototype Voice-Controlled Interactive Search Engine for Lifelog. In *Proceedings of the 5th Annual on Lifelog Search Challenge* (Newark, NJ, USA) *(LSC '22)*. Association for Computing Machinery, New York, NY, USA, 43–47. https://doi.org/10.1145/3512729.3533009
[4] Wei-Hong Ang, An-Zi Yen, Tai-Te Chu, Hen-Hsen Huang, and Hsin-Hsi Chen. 2021. LifeConcept: An Interactive Approach for Multimodal Lifelog Retrieval through Concept Recommendation *(LSC '21)*. Association for Computing Machinery, New York, NY, USA, 47–51. https://doi.org/10.1145/3463948.3469070
[5] Alexey Bochkovskiy, Chien-Yao Wang, and Hong-Yuan Mark Liao. 2020. YOLOv4: Optimal Speed and Accuracy of Object Detection. arXiv:2004.10934 [cs.CV]
[6] Alexander Faisst and Björn Jónsson. 2021. LifeMon: A MongoDB-Based Lifelog Retrieval Prototype. 75–80. https://doi.org/10.1145/3463948.3469066
[7] Cathal Gurrin, Hideo Joho, Frank Hopfgartner, Liting Zhou, and Rami Albatal. 2016. NTCIR Lifelog: The First Test Collection for Lifelog Research. *Proceedings of the 39th International ACM SIGIR conference on Research and Development in Information Retrieval - SIGIR '16* (2016), 705–708. https://doi.org/10.1145/2911451.2914680
[8] Cathal Gurrin, Hideo Joho, Frank Hopfgartner, Liting Zhou, and Rami Albatal. 2016. Overview of NTCIR-12 Lifelog Task. NTCIR.
[9] Cathal Gurrin, Hideo Joho, Frank Hopfgartner, Liting Zhou, Rashmi Gupta, Rami Albatal, Dang Nguyen, and Duc Tien. 2017. Overview of NTCIR-13 Lifelog-2 Task. NTCIR.
[10] 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.
[11] 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

[12] Cathal Gurrin, Klaus Schoeffmann, Hideo Joho, Andreas Leibetseder, Liting Zhou, Aaron Duane, Dang Nguyen, Duc Tien, Michael Riegler, Luca Piras, et al. 2019. Comparing approaches to interactive lifelog search at the lifelog search challenge (LSC2018). *ITE Transactions on Media Technology and Applications* 7, 2 (2019), 46–59.

[13] Zhiyu He. 2022. THUIR-LL at the NTCIR-16 Lifelog-4 Task: Enhanced Interactive Lifelog Search Engine. https://api.semanticscholar.org/CorpusID:251712653

[14] Silvan Heller, Ralph Gasser, Mahnaz Parian-Scherb, Sanja Popovic, Luca Rossetto, Loris Sauter, Florian Spiess, and Heiko Schuldt. 2021. Interactive Multimodal Lifelog Retrieval with Vitrivr at LSC 2021. In *Proceedings of the 4th Annual on Lifelog Search Challenge* (Taipei, Taiwan) *(LSC '21)*. Association for Computing Machinery, New York, NY, USA, 35–39. https://doi.org/10.1145/3463948.3469062

[15] Nhat Hoang-Xuan, Hoang-Phuc Trang-Trung, E.-Ro Nguyen, Thanh-Cong Le, Mai-Khiem Tran, Tu-Khiem Le, Van-Tu Ninh, Cathal Gurrin, and Minh-Triet Tran. 2022. Flexible Interactive Retrieval SysTem 3.0 for Visual Lifelog Exploration at LSC 2022. In *LSC@ICMR 2022: Proceedings of the 5th Annual on Lifelog Search Challenge, Newark, NJ, USA, June 27 - 30, 2022*, Cathal Gurrin, Graham Healy, Liting Zhou, Björn Þór Jónsson, Duc-Tien Dang-Nguyen, Jakub Lokoc, Minh-Triet Tran, Wolfgang Hürst, Luca Rossetto, and Klaus Schoeffmann (Eds.). ACM, 20–26. https://doi.org/10.1145/3512729.3533013

[16] Andreas Leibetseder and Klaus Schoeffmann. 2021. lifeXplore at the Lifelog Search Challenge 2021. 23–28. https://doi.org/10.1145/3463948.3469060

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

[18] Thao-Nhu Nguyen, Tu-Khiem Le, Van-Tu Ninh, Ly-Duyen Tran, Manh-Duy Nguyen, A. Caputo, Sinéad Smyth, Graham Healy, and Cathal Gurrin. [n. d.]. DCU and HCMUS at NTCIR-16 Lifelog-4. https://api.semanticscholar.org/CorpusID:251466185

[19] 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

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

[21] 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

[22] 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]

[23] 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. *CoRR* abs/2103.00020 (2021).

[24] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. 2022. High-Resolution Image Synthesis with Latent Diffusion Models. arXiv:2112.10752 [cs.CV]

[25] Luca Rossetto, Matthias Baumgartner, Ralph Gasser, Lucien Heitz, Ruijie Wang, and Abraham Bernstein. 2021. Exploring Graph-Querying Approaches in Life-Graph. In *Proceedings of the 4th Annual on Lifelog Search Challenge* (Taipei, Taiwan) *(LSC '21)*. Association for Computing Machinery, New York, NY, USA, 7–10. https://doi.org/10.1145/3463948.3469068

[26] Luca Rossetto, Ralph Gasser, Loris Sauter, Abraham Bernstein, and Heiko Schuldt. 2021. A System for Interactive Multimedia Retrieval Evaluations. In *MultiMedia Modeling*, Jakub Lokoč, Tomáš Skopal, Klaus Schoeffmann, Vasileios Mezaris, Xirong Li, Stefanos Vrochidis, and Ioannis Patras (Eds.). Springer International Publishing, Cham, 385–390.

[27] Jihye Shin, Alexandra Waldau, Aaron Duane, and Björn Jónsson. 2021. PhotoCube at the Lifelog Search Challenge 2021. 59–63. https://doi.org/10.1145/3463948.3469073

[28] 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. *Proceedings of the 4th Annual on Lifelog Search Challenge* (2021).

[29] 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

[30] Minh-Triet Tran, Thanh-An Nguyen, Quoc-Cuong Tran, Mai-Khiem Tran, Khanh Nguyen, Van-Tu Ninh, Tu-Khiem Le, Hoang-Phuc Trang-Trung, Hoang-Anh Le, Hai-Dang Nguyen, Trong-Le Do, Viet-Khoa Vo-Ho, and Cathal Gurrin. 2020. FIRST - Flexible Interactive Retrieval SysTem for Visual Lifelog Exploration at LSC 2020. In *Proceedings of the Third ACM Workshop on Lifelog Search Challenge, LSC@ICMR 2020, Dublin, Ireland, June 8-11, 2020*, Cathal Gurrin, Klaus Schöffmann, Björn Þór Jónsson, Duc-Tien Dang-Nguyen, Jakub Lokoc, Minh-Triet Tran, and Wolfgang Hürst (Eds.). ACM, 67–72. https://doi.org/10.1145/3379172.3391726

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

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