LifeInsight 2.0: An Enhanced Approach for Automated Lifelog Retrieval in LSC'24

Gia-Huy Vuong*
Van-Son Ho*
Tien-Thanh Nguyen-Dang
Xuan-Dang Thai
Minh-Triet Tran†
University of Science, VNU-HCM
Vietnam National University
Ho Chi Minh City, Vietnam
vghuy@selab.hcmus.edu.vn
{20120021,19120036,19120003}@student.hcmus.edu.vn
tmtriet@fit.hcmus.edu.vn

Thang-Long Nguyen-Ho
Minh-Khoi Pham
Tu-Khiem Le
Van-Tu Ninh
Graham Healy
Cathal Gurrin
Dublin City University
Dublin, Ireland
thanglong.nguyenho27@mail.dcu.ie
minhkhoi.pham@adaptcentre.ie
tukhiem.le4@mail.dcu.ie
tu.ninhvan@adaptcentre.ie
graham.healy@dcu.ie
cathal.gurrin@dcu.ie

ABSTRACT

We introduce the LifeInsight 2.0 system – an enhanced version of LifeInsight, built specifically for the sixth annual Lifelog Search Challenge (LSC'23). LifeInsight 2.0 leverages the core functionalities of LifeInsight while incorporating significant improvements to address performance bottlenecks. This refined architecture aims to deliver superior search capabilities within the LSC'24. LifeInsight 2.0 employs an ensemble approach combining two powerful foundation models: CLIP (Contrastive Language-Image Pretraining) and BLIP2 (Bootstrapping Language-Image Pretraining) model. In addition, the system incorporates a temporal query mechanism and an automatic query parser. The former enables LifeInsight 2.0 to interpret queries that include time-based information, while the latter specifically handles tasks involving question answering.

CCS CONCEPTS

• **Information systems** → **Multimedia databases**; **Users and interactive retrieval**; *Search interfaces*; • **Human-centered computing** → **Interactive systems and tools**.

KEYWORDS

lifelog, interactive retrieval, automatic retrieval, spatial insights, AI-based assistance

*These authors contributed equally to this research.
†Corresponding author

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

ACM ISBN 979-8-4007-0550-2/24/06
https://doi.org/10.1145/3643489.3661112

ACM Reference Format:
Gia-Huy Vuong, Van-Son Ho, Tien-Thanh Nguyen-Dang, Xuan-Dang Thai, Minh-Triet Tran, Thang-Long Nguyen-Ho, Minh-Khoi Pham, Tu-Khiem Le, Van-Tu Ninh, Graham Healy, and Cathal Gurrin. 2024. LifeInsight 2.0: An Enhanced Approach for Automated Lifelog Retrieval in 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/10.1145/3643489.3661112

1 INTRODUCTION

Technological advancements, including wearable devices, cameras, location-tracking systems, and smartphones, have enabled us to record our life experiences continuously. This process of constant self-documentation is referred to as lifelogging. Lifelogging generates a substantial personal multimedia repository, termed lifelog data, which encapsulates diverse facets of our everyday lives [7]. This data can encompass photos collected passively, GPS records, and even physiological data. The release of a public lifelog dataset through the Lifelog Search Challenge (LSC) has opened doors for various research efforts in the field. This data empowers researchers to explore how lifelogging can be harnessed to benefit users' daily lives. A key area of exploration, which is also the main focus of the LSC, is the development of an "auxiliary memory assistant." This intelligent system aims to act as a real-time search engine for users' vast personal archives, allowing them to retrieve specific memories instantly [8].

Lifelog data often involves large quantities of passively captured images, lacking captions or labels beyond their metadata. This metadata typically includes timestamps, locations, identified objects within the image, and any text detected in the scene. Early lifelog retrieval systems, like LifeSeeker [14], lifeXplore [23], and FIRST [11], relied heavily on metadata search through Database Management Systems (DBMS) or Elastic Search. This limited their ability to understand the meaning of the content. Nevertheless, Vision-language models like CLIP [20] and BLIP [13] have opened exciting possibilities in lifelog retrieval. These models enable semantic search, allowing users to describe the desired content without needing exact keywords from the metadata. This eliminates the need for complex query parsing and leads to significantly improved search accuracy.

This paper presents LifeInsight 2.0 – a new iteration of the LifeInsight system [17], specifically designed for the LSC'24 challenge [8]. The key difference between the new and previous versions of LifeInsight lies in adopting an ensemble technique. This approach, demonstrably effective in enhancing AI system accuracy, allows LifeInsight 2.0 to leverage the combined strengths of BLIP [13] and CLIP [20] models for improved image-text retrieval, replacing the prior version's reliance solely on BLIP. Furthermore, we introduce a temporal query mechanism that empowers users to formulate queries incorporating temporal information. Additionally, we provide an automatic query parser that streamlines user queries, improving the system's returned results. The concept behind this feature draws inspiration from our previous work in the NTCIR-17 conference [18]. However, to further enhance our approach, we have decided to use the more novel model (detailed in Subsection 4.4) for extracting the main context while keeping the Stanza model [19] to handle named entity recognition.

In summary, we adapted the comprehensive lifelog retrieval system LifeInsight 2.0 which was specially developed for the Seventh Annual ACM Lifelog Search Challenge [8] by adding the following enhancement:

(1) We cluster the lifelog moments into events based on the context of the lifelog images, thereby reducing the number of embeddings inserted into our vector database (Milvus [1] and Elastic [2], increasing the searching speed of the LifeInsight 2.0 system.
(2) This paper explores the difference between utilizing an ensemble technique combining CLIP [20] and BLIP [13] models and solely relying on BLIP for image-text search.
(3) LifeInsight 2.0 utilizes large language models (LLMs) to construct an automated query parser. This parser serves as a preprocessing step, extracting conceptual information like location, context, or time from user queries, thereby augmenting its search capabilities.
(4) LifeInsight 2.0 empowers users with a user-friendly temporal query mechanism, enabling them to formulate queries incorporating temporal information and search for relevant results.

2 RELATED RESEARCH

The Lifelog Search Challenge (LSC) has gained popularity over the years, attracting more participants from various organizations. This challenge focuses on creating interactive retrieval systems. These systems aim to find specific images from a large collection representing a person's life events, all within a limited timeframe and based on a user's query.

Various systems, including Memoria [21], LifeSeeker [16], vitrivr [9], FIRST [29], and Myscéal [27], have offered multiple search modalities based on concepts. Lifegraph [22] and LifeConcept [3] utilized knowledge graphs and concept recommendation methods like ConceptNet to facilitate retrieval by linking relevant concepts with images. Other systems such as lifeXplore [24], PhotoCube [26], and LifeMon [5] employed convolutional neural networks (CNNs) like YOLOv4 [4] and traditional object detectors for content analysis. These systems primarily used Database Management Systems (DBMS) or Elasticsearch data retrieval mechanisms to align user queries with visual concepts and metadata effectively. A number of systems, including LifeSeeker 4.0 [15], E-Myscéal [28], Memento 2.0 [1], FIRST 3.0 [12], and Voxento [2], incorporated vision-language pre-trained models, specifically the CLIP model [20]. These systems demonstrated significant performance improvements in zero-shot image-text retrieval compared to their previous versions.

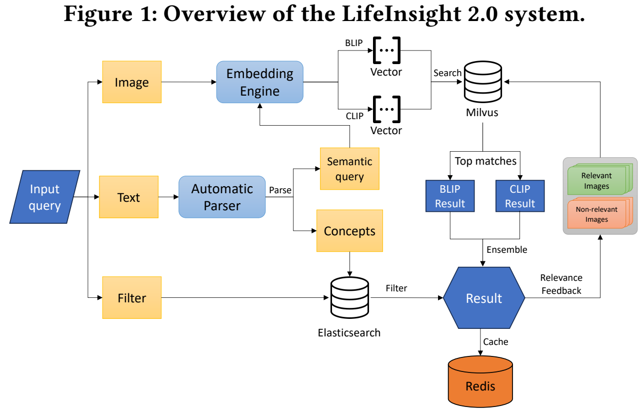
Semantic-driven systems strive to extract insights not only from visual content but also from the semantic context of the query description, leading to more accurate results. For the LSC'23, we created LifeInsight, a semantic-driven lifelog retrieval system. This system leverages the BLIP vision-language pre-training model [13] and incorporates various AI features to support the retrieval process alongside the semantic search function.

3 OVERVIEW OF LIFEINSIGHT 2.0

3.1 System Overview

The overview of LifeInsight 2.0 is depicted in Figure 1, demonstrating the improvements over the previous version. Specifically, LifeInsight 2.0 can automatically transform an input query into a semantic query, which includes context and concepts that aid in searching within Elasticsearch. Furthermore, LifeInsight 2.0 incorporates multiple encoders, in this case, BLIP-2 [13] and CLIP [20], to boost performance, as opposed to relying solely on BLIP-2 as in the case of LifeInsight.

Figure 1: Overview of the LifeInsight 2.0 system.



3.2 User Interface

LifeInsight 2.0's main user interface builds upon the familiar design of LifeInsight [17] while offering enhanced functionality for a more intuitive user experience. In response to user feedback regarding convenience, LifeInsight 2.0 positions the side tab bar on the left side, a departure from the previous version's right-side placement. LifeInsight 2.0 retains the core chat interface experience but streamlines the user experience by strategically adding and

[1] https://milvus.io/
[2] https://www.elastic.co/

Figure 2: Relevant images displayed in LifeInsight 2.0

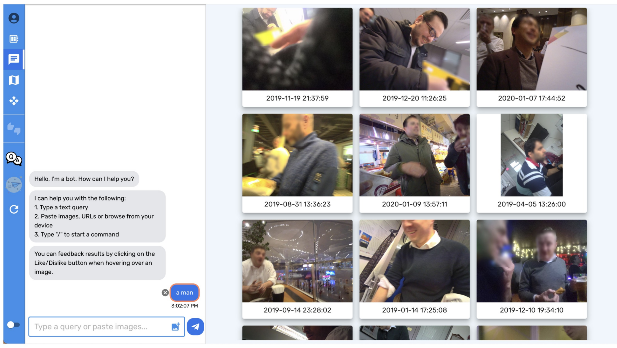
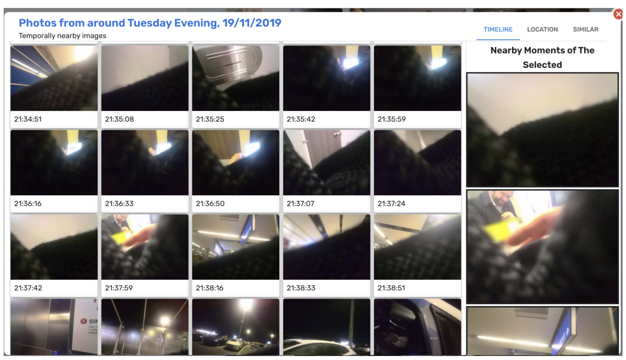


Figure 3: Detail information displayed in LifeInsight 2.0



removing buttons. In Figure 2, we showcase the key functionalities of our LifeInsight 2.0 system. The right section of the user interface displays retrieval results in a vertically scrollable panel.

Selecting an image from the retrieval results triggers a pop-up window (as seen in Figure 3) with three distinct tabs. Each tab provides a unique perspective on the chosen image collection, including:

- **Timeline**: This tab displays temporally nearby images to the selected image. Additionally, a vertical panel appears on the right side of the pop-up window, showcasing enlarged images from the previous and next moments.
- **Location**: This tab arranges all images taken at the same location as the selected image by similarity score. This allows the user to quickly find pictures taken at the same place as the selected one and are visually similar to it.
- **Visual Similarity**: This tab displays visually identical images to the selected image, with an option to filter visually similar photos captured at the same location. This feature would enable users to find images with similar visual content, even if they were not taken at the same time or place as the selected image.

4 MAIN COMPONENTS OF LIFEINSIGHT 2.0

4.1 Data Preprocessing

While LifeInsight 2.0 leverages data preprocessing techniques like LifeInsight [17], we achieve superior performance by representing each daily event as a single, compact vector instead of using all individual event vectors. This is accomplished through an event segmentation task (detailed in 4.2), which efficiently partitions all daily images into distinct event categories. Our approach significantly reduces the number of vectors and metadata stored in Milvus and Elasticsearch, leading to dramatically improved search speed. Additionally, it minimizes duplicate images, enabling users to find relevant information more efficiently.

4.2 Event Segmentation

We employ a vector clustering approach to categorize images from a day into distinct event groups. Each image is represented by a feature vector. Clustering is performed based on the cosine similarity between these vectors. We used a sequential process, where the similarity score between a given vector and the centroid (average) of each existing cluster is calculated. The vector is assigned to that cluster if the score exceeds a predefined threshold. Otherwise, it is compared to the centroids of other clusters. A new cluster is created with the given vector as its initial member if no suitable cluster is found.

4.3 Semantic Search

Unlike LifeInsight, which relied solely on Bootstrapping Language-Image Pre-training (BLIP) [13], our system utilizes an ensemble combining BLIP and Contrastive Language-Image Pre-training (CLIP) [20]. This approach mitigates potential biases and enhances performance. For instance, a given query q will be encoded to a vector $\vec{q}_b$ (output when q is fed into BLIP [13]), and a vector $\vec{q}_c$ (output when q is fed into CLIP [20]). The final input embedding vector is formulated as follows:

$$\vec{q}_{\text{avg}} = \beta_c \vec{q}_c + \beta_b \vec{q}_b$$

where β_b (0.6 in LifeInsight 2.0) and β_c (0.4 in LifeInsight 2.0) are refined weights.

Similar to LifeInsight, the computation of image similarity, resulting in a ranked list relevant to a specific query description, is performed by Milvus [3]. The Inner Product compares the distance between embeddings in the vector space.

4.4 Automatic Query Parser

Unlike our previous system, LifeInsight [17], which relied solely on raw user queries for semantic search, the enhanced LifeInsight 2.0 leverages an automatic query parser to extract key concepts like main context, location, date, and time. This empowers LifeInsight 2.0 to understand user intent more effectively. The automatic query parser in LifeInsight 2.0 utilizes two key components: the Python Stanford NLP library (Stanza) [19] for named entity recognition (NER) and the GEMMA-2B model [6] for parsing the main context within a query. This enables LifeInsight 2.0 to excel at automatic

[3]https://milvus.io/

retrieval tasks and question-answering tasks. To illustrate the power of automatic query parsing, let us consider some examples:

RAW INPUT: *I was praying to small golden Buddha in a tunnel. There were plants and offerings around the Buddha. It was inside of a tourist park with a large ornamental tower. It was in September 2019 in Thailand.* Following automatic query parsing within LifeInsight 2.0 the output will be:

Location:	Thailand.
Date:	September 2019.
Concepts:	park, tower.
Main Context:	Praying to a small golden Buddha in a tunnel. There were offerings around the Buddha. Inside of a tourist park with a large ornamental tower. It was in September 2019.

RAW INPUT: *Moments in which the lifelogger was praying to a small golden Buddha in a tunnel. There were plants and offerings around the Buddha. It was inside of a tourist park with a large ornamental tower.* Following automatic query parsing within LifeInsight 2.0 the output will be:

Concepts:	park, tower.
Main Context:	Praying to a small golden Buddha in a tunnel. There were offerings around the Buddha. Inside of a tourist park with a large ornamental tower.

RAW INPUT: *I think it was the second time I visited the house with the stone shed/hovel. The shed was under green trees on a beautiful sunny day. It takes 2 hours to drive there and two hours to drive back home. It was in the middle of Ireland on the 29th April 2020.* Following automatic query parsing within LifeInsight 2.0, the output will be:

Date:	the 29th April 2020.
Location:	Ireland.
Time:	2 hours, two hours.
Ordinal:	second.
Concepts:	time, house, stone, green, sunny, day.
Main Context:	Visiting the house with the stone shed/hovel. The shed was under green trees. Driving to a place and driving back home. It was in the middle of Ireland.

RAW INPUT: *After a short relaxing walk, I reached the edge of a lake. There were mountains and trees, but very few people. It was a cold day in Spring in Wicklow in 2019.* Following automatic query parsing within LifeInsight 2.0 the output will be:

Date:	Spring, 2019.
Location:	Wicklow.
Concepts:	lake, people, day.
Main Context:	After a short relaxing walk. Mountains and trees. It was a cold day in Spring in Wicklow in 2019.

4.5 Temporal query mechanism

The power of the temporal query mechanism, which has been proven over the years, is particularly evident in the VITRIVR system. A paper titled "Multi-Stage Queries and Temporal Scoring in VITRIVR" was published by Silvan Heller and his team, further highlighting its effectiveness [10]. This technique's power was also showcased in the ViewsInsight system [30], a platform specifically engineered for the Video Browser Showdown 2024 (VBS2024) [25]. Recognizing the mechanism's proficiency in managing temporal information during searches, we have seamlessly integrated it into our system, LifeInsight 2.0, to address the challenges presented by LSC'24 [8].

This mechanism relies on three user inputs which consist of the main query and two temporal descriptions. The temporal descriptions provide details about events occurring before and after the event mentioned in the query. Within the system, the algorithm extracts relevant information from these queries. Subsequently, it updates the scores of the "now" query by considering the highest score from the "before" and "after" queries associated with the same day. Finally, the algorithm generates a list of search results, organized based on their newly updated scores.

The pseudo-code of this process is detailed in Algorithm 1. This powerful feature is seamlessly integrated into the system, enabling it to retrieve relevant results effectively. This makes the system more efficient and user-friendly

5 SOME USAGE SCENARIOS

Query 1: *I was praying to small golden Buddha in a tunnel. There were plants and offerings around the Buddha. It was inside of a tourist park with a large ornamental tower. It was in September 2019 in Thailand.*

When you enter a query like "praying golden Buddha in Thailand 2019," our system, LifeInsight 2.0, can automatically break it down. It identifies the main context (praying golden Buddha) and refines the search by extracting concepts like location (Thailand) and date (2019). This process is illustrated in Figure 4.

Query 2: *It was an outdoor outdoor kitchen. I remembered it was BBQ party happening in 2020.*

This query can be easily addressed using the visual search feature in LifeInsight 2.0. Simply download an image of an "outdoor BBQ kitchen" and upload it to LifeInsight 2.0's search bar. The system will then identify similar outdoor BBQ kitchen designs, providing the most relevant results based on the image (shown in Fig 5).

Query 3: *After a short relaxing walk, I reached the edge of a lake. There were mountains.*

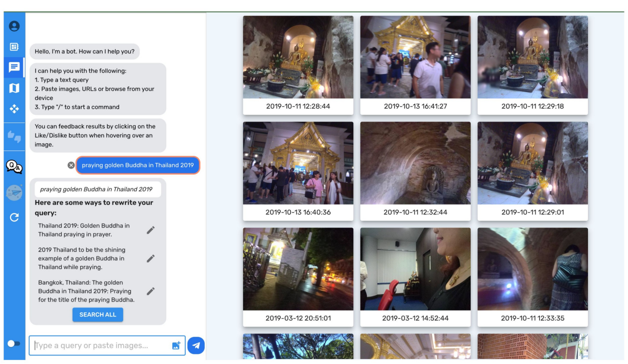
With LifeInsight 2.0's temporal query mechanism, you can do just that. Simply enter "walking > reach the edge of a lake and

Algorithm 1 Temporal Search Algorithm

Require: a, b, c, $range$

// The variables a, b, and c represent a before, now, and after query, respectively.
// The variable $range$ represents the maximum duration when before, now, and after events occur.
$A \leftarrow$ Dictionary containing information retrieved from the system when a (now query) is searched for.
$B \leftarrow$ Dictionary containing information retrieved from the system when b (now query) is searched for.
$C \leftarrow$ Dictionary containing information retrieved from the system when c (now query) is searched for.
for all b_i in B **do**
  $A_1 \leftarrow \{a_j \in A_1 \mid a_j.date\_id = b_i.date\_id \wedge a_j.timestamp \in (b_i.timestamp - range,\ b_i.timestamp)\}$
  $C_1 \leftarrow \{c_j \in C_1 \mid c_j.date\_id = b_i.date\_id \wedge c_j.timestamp \in (b_i.timestamp,\ b_i.timestamp + range)\}$
  if A_1 is not empty **then**
    $max\_score\_a \leftarrow$ Highest score in A_1.
  end if
  if C_1 is not empty **then**
    $max\_score\_c \leftarrow$ Highest score in C_1.
  end if
  b_i.score $\leftarrow b_i$.score + max_score_a + max_score_c
end for
Sort B in descending order based on updated scores
return Sorted B as relevant image results

Figure 4: Utilize visual search to identify the matched frames.



mountains" (described in Fig 6) to search for images that transition from walking to a view of a lake and mountains. In this example, "walking" is the "before" event, and "reach the edge of a lake and mountains" is the "after" event.

6 CONCLUSION

In conclusion, the LifeInsight 2.0 lifelog retrieval system is a comprehensive system that employs several mechanisms and features to provide insightful and relevant search results for lifelog data. By incorporating semantic search mechanisms from state-of-the-art

Figure 5: Utilize visual search to identify the matched frames.

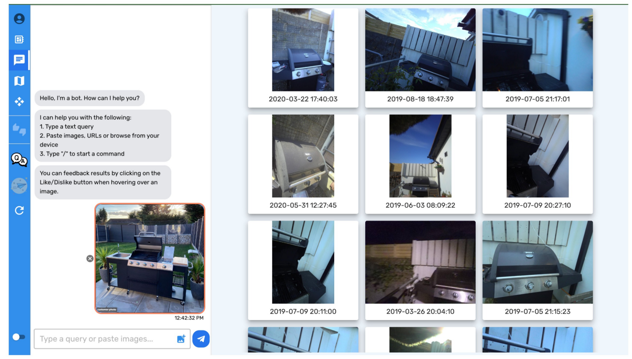
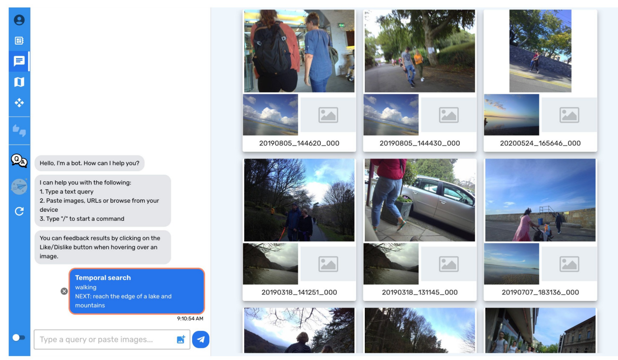


Figure 6: Utilize visual search to identify the matched frames.



systems and focusing on using spatial information, LifeInsight 2.0 provides an effective approach for retrieving relevant information from lifelog data. The use of the Bootstrapping Language-Image Pre-training (BLIP) model for zero-shot image-text retrieval and Elastic Search for filtering irrelevant images would enhance the precision and recall scores of the system based on previous experiments from other related systems with the same search mechanism. The integration of visual similarity search functionality and explicit relevance feedback enables the system to provide more accurate search results, while AI-based query description rewriting and visual example generation features further support end-users during the retrieval process.

ACKNOWLEDGEMENT

Acknowledgement: This publication has emanated from research supported in part by research grants from Science Foundation Ireland under grant numbers SFI/13/RC/2106_P2, SFI/12/RC/2289_P2 and was funded by Vingroup Innovation Foundation (VINIF) under project code VINIF.2019.DA19.

REFERENCES

[1] 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

[2] 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

[3] 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

[4] Alexey Bochkovskiy, Chien-Yao Wang, and Hong-Yuan Mark Liao. 2020. YOLOv4: Optimal Speed and Accuracy of Object Detection. arXiv:cs.CV/2004.10934

[5] Alexander Faisst and Björn Jónsson. 2021. LifeMon: A MongoDB-Based Lifelog Retrieval Prototype. 75–80. https://doi.org/10.1145/3463948.3469066

[6] Thomas Mesnard Gemma Team, Cassidy Hardin, Robert Dadashi, Surya Bhupatiraju, Laurent Sifre, Morgane Rivière, Mihir Sanjay Kale, Juliette Love, Pouya Tafti, Léonard Hussenot, and et al. 2024. Gemma. (2024). https://doi.org/10.34740/KAGGLE/M/3301

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

[8] 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

[9] 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

[10] Silvan Heller, Loris Sauter, Heiko Schuldt, and Luca Rossetto. 2020. Multi-Stage Queries and Temporal Scoring in Vitrivr. 1–5. https://doi.org/10.1109/ICMEW46912.2020.9105954

[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] 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

[13] Junnan Li, Dongxu Li, Caiming Xiong, and Steven C. H. Hoi. 2022. BLIP: Bootstrapping Language-Image Pre-training for Unified Vision-Language Understanding and Generation. *CoRR* abs/2201.12086 (2022).

[14] 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

[15] 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

[16] Thao-Nhu Nguyen, Tu-Khiem Le, Van-Tu Ninh, Minh-Triet Tran, Nguyen Thanh Binh, Graham Healy, Annalina Caputo, and Cathal Gurrin. 2021. LifeSeeker 3.0: An Interactive Lifelog Search Engine for LSC'21. In *Proceedings of the 4th Annual on Lifelog Search Challenge* (Taipei, Taiwan) *(LSC '21)*. Association for Computing Machinery, New York, NY, USA, 41–46. https://doi.org/10.1145/3463948.3469065

[17] 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

[18] 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. As the demand for personalized data retrieval systems continues to grow, recent research has emphasized the development of lifelog retrieval mechanisms. Many new research and methods have focused on studying the integration of user interactions and feedback into search engines. In this paper, we introduce the automation approach of LifeInsight, a retrieval system designed explicitly for the NTCIR-17 Lifelog-5 Automatic Task, facilitating a seamless search experience and efficient data mining. Our method entails a two-fold process, where we first enrich the metadata from the raw query, followed by the composition of the retrieval method from input entities. Our proposed system not only enhances the search process but also ensures a comprehensive and detailed analysis of lifelog data for diverse applications. By focusing primarily on the automatic sub-task, we demonstrate the efficacy of our LifeInsight retrieval algorithm, showcasing competitive results that rival those of an expert user.

[19] Peng Qi, Yuhao Zhang, Yuhui Zhang, Jason Bolton, and Christopher D. Manning. 2020. Stanza: A Python Natural Language Processing Toolkit for Many Human Languages. *CoRR* abs/2003.07082 (2020).

[20] 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). arXiv:2103.00020 https://arxiv.org/abs/2103.00020

[21] Ricardo Ribiero, Alina Trifan, and Antonio J. R. Neves. 2022. MEMORIA: A Memory Enhancement and MOment RetrIeval Application for LSC 2022. In *Proceedings of the 5th Annual on Lifelog Search Challenge* (Newark, NJ, USA) *(LSC '22)*. Association for Computing Machinery, New York, NY, USA, 8–13. https://doi.org/10.1145/3512729.3533011

[22] Luca Rossetto, Matthias Baumgartner, Ralph Gasser, Lucien Heitz, Ruijie Wang, and Abraham Bernstein. 2021. Exploring Graph-Querying Approaches in LifeGraph. 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

[23] 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

[24] 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

[25] Klaus Schoeffmann, Jakub Lokoc, and Werner Bailer. 2020. 10 years of video browser showdown. In *MMAsia 2020: ACM Multimedia Asia, Virtual Event / Singapore, 7-9 March, 2021*, Tat-Seng Chua, Jingdong Wang, Qi Tian, Cathal Gurrin, Jia Jia, Hanwang Zhang, and Qianru Sun (Eds.). ACM, 73:1–73:3. https://doi.org/10.1145/3444685.3450215

[26] 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

[27] 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).

[28] 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

[29] 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

[30] Gia-Huy Vuong, Van-Son Ho, Tien-Thanh Nguyen-Dang, Xuan-Dang Thai, Tu-Khiem Le, Minh-Khoi Pham, Tu Ninh, Cathal Gurrin, and Minh-Triet Tran. 2024. *ViewsInsight: Enhancing Video Retrieval for VBS 2024 with a User-Friendly Interaction Mechanism.* 400–406. https://doi.org/10.1007/978-3-031-53302-0_38