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/10.20736/0002001329

Overview of the NTCIR-17 Lifelog-5 Task

Liting Zhou
Dublin City University
Ireland

Graham Healy
Dublin City University
Ireland

Cathal Gurrin
Dublin City University
Ireland

Ly Duyen Tran
Dublin City University
Ireland

Naushad Alam
Dublin City University
Ireland

Hideo Joho
Tsukuba University
Japan

Longyue Wang
AI Lab in Tencent
China

Tianbo Ji
Nantong University
China

Chenyang Lyu
Mohamed bin Zayed University of
Artificial Intelligence
Abu Dhabi

Duc-Tien Dang-Nguyen
University of Bergen & Kristiania
University College
Norway

ABSTRACT

NTCIR-17 witnessed the fifth iteration of the Lifelog task, which was designed to facilitate the comparative evaluation of various approaches for automatic and interactive information retrieval from multimodal lifelog archives. Within this paper, we elucidate the utilization of the test collection, delineate the specified tasks, provide an overview of the submissions, and present the findings derived from the NTCIR17 Lifelog-5 LSAT sub-task. Our conclusion includes recommendations for potential future developments in the realm of lifelog tasks.

KEYWORDS

lifelog, information retrieval, quantified self, personal data

SUBTASKS

LSAT subtask (English), LIT subtask, LQAT subtask

1 INTRODUCTION

NTCIR-17 [12] hosted the fifth edition of the Lifelog task. The aim of the lifelog task is to foster comparative benchmarking of approaches to automatic and interactive information retrieval from multimodal lifelog archives. In this edition of the Lifelog task, we focused on three subtasks: the Lifelog Semantic Access (sub)Task (LEST), which is a conventional ad-hoc retrieval task for lifelogs; Lifelog Insight (sub)Task, which is for exploring knowledge mining and visualisation of lifelogs by setting general challenges for the participants to address; Lifelog question answer (sub)task(LQAT), which is used for lifelog quetsion answer task.The LEST task had been central to the previous NTCIR Lifelog tasks at NTCIR-12 [6], NTCIR-13 [7], NTCIR-14 [8] and NTCIR16 [17]. The LIT tasks were introduced during NTCIR12 and subsequently reintroduced during NTCIR17.

Dodge and Kitchin introduced the concept of lifelogs in their work [2]. A lifelog typically comprises a diverse range of data types, including image and video content captured by wearable cameras like Sensecam and Narrative, audio content from personal audio devices, biometric sensor data from activity trackers (e.g., wristbands or phones as mentioned in [1]), and information from the media consumed by the lifelogger, among other sources.

Early retrieval systems for lifelog data, such as the MyLifeBits system [5] and the Sensecam Browser [3], were primarily browsing engines relying on databases for access. However, subsequent research revealed that a faceted-multimodal search engine, even a simple one, is significantly faster and more effective at finding known items within extensive lifelogs [4]. Despite this, there were few search engines specifically designed for lifelog data, and there was no means of comparing their effectiveness. This deficiency served as one of the motivations for organizing the challenge at the NTCIR conference, which supports a comparative evaluation of lifelog data storage and retrieval approaches.

In this paper, we contribute by introducing the task and providing an overview of the performance of participating teams. The remaining sections of the paper detail the dataset used, the three subtasks, the topics explored, the comparisons among the participants, and our thoughts on the future of lifelog benchmarking.

2 DATASET DESCRIPTION

2.1 Overview

NTCIR-17-Lifelog-5 reuses an existing dataset, the LSC'22 dataset [11], which is a multimodal dataset that is four months in size, from one active lifelogger. The dataset consists of four files, which were made available to participants who signed up for the lifelog task and who agreed to the terms of access. The data continued within these files was gathered using multiple wearable sensors, such as PoV cameras, biometric smartwatches and location and activity loggers on a smartphone. The data gathering phases occurred in 2019 (twelve months) and 2022 (six months), giving a total of 16 months of lifelog data, which was gathered 24 x 7 and organised into minutes in an XML form. For dataset examples, see [10]. The

UTC timestamp was the used as the alignment factor for these data sources. The four files that comprise the dataset were:

- Metadata for the collection (112.3MB), consisting of textual XML metadata representing time, physical activities, biometrics and locations.
- Core Image Dataset (48GB, Comprared) of 725,950 wearable camera images, fully redacted and anonymised in 1024 x 768 resolution, captured using a Narrative Clip device. These images were collected during 2019-2020 and captured during regular waking hours by the same one individual. All faces and readable text have been removed, as well as certain scenes and activities manually filtered out (by the data gatherer / lifelogger) to respect privacy expectations.
- Visual Concepts (121MB) extracted from the non-redacted version of the visual dataset. The Visual Concepts data file includes detected scenes and concepts for each image (processed over the non-redacted version of the images). the objects detected automatically from the image. We use the object category list of 2014-2017 COCO datasets [14] with 80 labels for annotation.
- Additional Data: MySceál's team[16] offers a supplementary metadata file which includes semantic location names and enhancements to the raw location data. Additionally, Voxento team has provided custom location metadata that addresses irregularities related to flights, representing flight locations as departing airport to arrival airport.

3 LIFELOG TASKS OVERVIEW

In the current year, we organized three distinct subtasks: Among these, only the LSAT subtask received four submissions. Regrettably, neither the LIT nor the LQAT subtask garnered any submissions. The main reason is the novelty of the latter two subtasks and limited dissemination of information regarding their existence with the broader research community:

- **The Lifelog Semantic Access Task (LSAT)** is a known-item search task that can be undertaken in an interactive or automatic manner. In this sub task, the participants have to retrieve a number of specific moments in a lifelogger's life. We define moments as semantic events, or activities that happened throughout the day. The task can best be compared to a known-item search task.
- **Lifelog Insight subTask (LIT)** The LIT task was exploratory in nature and the aim of this subtask was to gain insights into the lifelogger's daily life activities. It followed the idea of the Quantified Self movement that focuses on the visualization of knowledge mined from self-tracking data to provide "self-knowledge through numbers". Participants were requested to provide insights about the lifelog data that support the lifelogger in the act of reflecting upon the data, facilitate filtering and provide for efficient/effective means of visualisation of the data. The LIT task included ten information needs representing the idea that one would use a lifelog as a source for self-reflection. We did not intend to have an explicit evaluation for this task, rather we expected all participants to being their demonstrations or reflective output at the NTCIR conference.
- **Lifelog Question Answer subTask (LQAT)** is to encourage comparative progress on the important Q&A topic from lifelogs. For this subtask, an augmented 85-day lifelog collection with over 15,000 multiple-choice questions and baseline will be provided, and participants can train and compare their lifelog QA models.

4 EVALUATION TASK DETAILS

As previously methioned, the NTCIR-17 lifelog task was structured around three sub-tasks. However, there is no participants engaged in the LIT and LQAT sub-tasks. Consequently, our primary focus in this context will be on evaluationg the LSAT sub-task, which has consistently been the most popular sub-task in all previous iterations of the lofelog task. The other tasks that typically explore event segmentation, multimodal annotation, and insight generation were not included in NTCIR-17.

4.1 Lifelog Semantic Access sub-Task (LSAT)

The LSAT sub-task presented participants with the challenging task of navigating 41 distinct topics using a lifelog retrieval system, with the ultimate goal of producing ranked results for evaluation. Participants had the flexibility to engage with this task in either an interactive or automatic manner.

In the case of automatic runs, the presumption was that the search process operated independently of any user interaction after the initial query construction phase. This non-interactive approach allowed participants to focus on developing effective query strategies, with no constraints on the time it took to execute their automatic runs. The primary objective of these runs was to facilitate a comparative analysis of various backend ranking algorithms, and we received submissions from four different systems for this category.

On the other hand, interactive runs assumed active user involvement in every aspect of the search process, encompassing query generation, refinement, and the selection of images deemed relevant for each topic. This user engagement could occur as a single phase or involve multiple stages, including relevance feedback and query reformulation. In the context of interactive runs, a time limit of 300 seconds was imposed for each topic to ensure a reasonable and consistent evaluation framework. To facilitate performance comparisons at different time points within the 300-second window, participants were encouraged to include a seconds-elapsed indicator in their submissions. This allowed us to gain insights into system behavior at various time cutoffs, ranging from instantaneous to the full 300 seconds, offering a comprehensive view of the effectiveness of interactive search strategies.

4.2 Topics

The LSAT sub-task entailed a comprehensive collection of 41 topics, thoughtfully structured in adherence to the well-established TREC format. Each topic encompassed integral components such as ID, Query, Description, and Narrative, as eloquently exemplified in Listing 1 and Listing 2. This meticulous structuring was undertaken to ensure conformity with the established standards of previous NTCIR-Lifelog topics. Each topic was meticulously characterized by its type, which could be ad-hoc or known-item, and was attributed

Table 1: Ad-Hoc Topics

ID	Title	Num-Rel
17001	Eating Avocado	208
17002	Taking Medication	17
17003	repairing electric shower	9
17004	Reading menu	1125
17005	Car stopping	1678
17006	Grocery shopping	738
17007	ATM	91
17008	'Bee Happy' t-shirt	85
17009	Sunset	8
17010	Flowers in window	27
17011	Drinking Guinness	2302
17012	Exotic birds	65
17013	Listening Music	138
17014	I like cake	82
17015	Buying whisky	30
17016	Trying on clothes	14
17017	Early flights	17

Table 2: Known-Item Topics

ID	Title	Num-Rel
17018	praying to small golden Buddha	8
17019	Visiting shed / hovel	4
17020	buying beans	1
17021	Mother Mary prays for us	1
17022	Lake	3
17023	Mug for sale	1
17024	Greek wine on a Sunday	45
17025	Cottage and shed for sale	12
17026	Two vinyl LPs (records)	3
17027	Meeting friends outside the Brazen Head	6
17028	Get back' on the roof	7
17029	Zombies on the platform?	11
17030	Preaching to a full room	1
17031	Buying hand soaps	8
17032	Airport pick up	34
17033	Oyster	3
17034	Man wearing yellow hat	222
17035	Eating sandwiches	101
17036	A man walking his dog	81
17037	Having lunch with Dermot	131
17038	Eye test	18
17039	Model train	5
17040	Man with Pink t-shirt	1
17041	Drinks on top of the Bangkok	155

to a specific user (uid). In this particular dataset, the user representation was consistently denoted as u1, owing to the single-user context.

The description segment of each topic played a pivotal role in equipping the user with essential information pertaining to the subject matter at hand, while the narrative component assumed a crucial role in disambiguating the criteria for relevance. This distinction was especially significant for interactive runs, where user guidance was vital.

In its entirety, the LSAT sub-task featured a carefully curated set of 41 topics. These topics comprised 17 ad-hoc queries and 24 known-item queries. The ad-hoc queries closely resembled traditional text retrieval queries, with the primary objective of unearthing a maximal number of pertinent items in response to a given query. The pursuit of these relevant items led participants through various events, making the search task an amalgamation of intrigue and complexity. Table 1 serves as a repository of the ad-hoc topics, with the topic ID initiating at 17001. The Num-Rel column in the table quantifies the count of relevant items as determined by pooled relevance judgments for each topic. These statistics offer invaluable insights into the effectiveness of the retrieval process.

The 24 known-item topics focused on solving targeted information needs that were typically solved by a single event with a small number of relevant images. Known-item topics are designed to simulate the human memory process of finding or remembering a specific event or activity (e.g. solving a task or attending a certain location). Table 2 shows the 24 known-item topics along with the number of relevant items in the relevance judgements. It is worth noting that the known-item topics are based on existing topics from the LSC'22 and LSC'23 [9] benchmarking workshop and as such, would be familiar to any participant who took part in the LSC'22 and LSC'23 exercise.

4.3 Relevance Judgements

Relevance judgments were annotated by professionals deeply entrenched in the realm of lifelog research. This curation involved the identification and assimilation of images considered relevant, which were then seamlessly integrated into the corpus of relevance judgments. To ensure the highest quality and precision, these judgments underwent a rigorous double review process, receiving meticulous scrutiny from the original lifelogger (u1).

For known-item topics, a comprehensive approach was taken to identify additional relevant items. Leveraging the knowledge and insights shared in the LSC'22 and LSC'23 workshops, where applicable, thoughtful efforts were made to incorporate these valuable findings into the collection of relevance judgments. This rigorous and thoughtful approach aimed to enhance the precision and comprehensiveness of the relevance judgments.

5 EVALUATION RESULTS

Considering the dichotomy of submissions within the LSAT subtask, we shall conduct an independent examination of each. Within the LSAT task, five participants made submissions, warranting a distinct discussion of the automatic and interactive facets. Subsequent sections will comprehensively dissect and present the assessment outcomes of the participating teams.

Listing 1: Ad-Hoc Topic 17001

```
<topic>
<id>17001</id>
<type>adhoc</type>
<uid>u1</uid>
<title>Eating Avocado</title>
<description>Find examples of when I
    was eating avocado for breakfast.
    </description>
<narrative>To be relevant, the images
    must show the lifelogger was eating
     avocado. Other fruits are not
    relevant</narrative>
</topic>
```

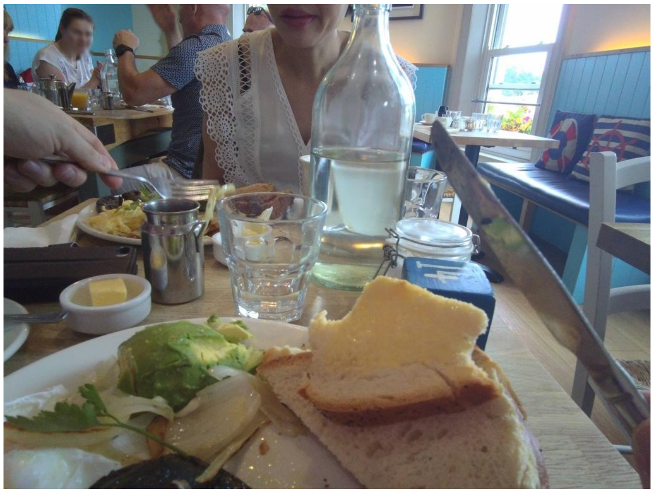

Figure 1: Example Relevant Image for Adhoc Topic 17001 (Eating Avocado)

Listing 2: Known-item Topic 16033

```
<topic>
<id>17025</id>
<type>knownitem</type>
<uid>u1</uid>
<title>Cottage and shed for sale</title
    >
<description>Find the moment when the
    lifelogger was viewing cottage and
    shed for sale</description>
<narrative>Relevant moments must show
    cottage and shed for sale. A stone
    cottage with a broken roof on the
    left side. The cottage was
    surrounded by tall trees and had 4
    windows and a door with yellow
    surrounds. </narrative>
</topic>
```

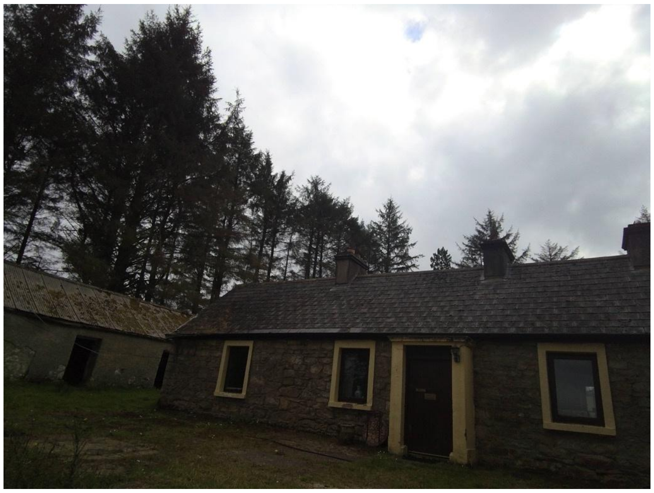

Figure 2: Example Relevant Image for Known-Item Topic 16025 (Cottage and shed for sale)

5.1 LSAT-Automatic Runs

Three distinct participants submitted automatic runs and four different systems were employed:

DCU MemoriEase: This team's approach involved the submission of an automatic query system, leveraging the power of the BLIP-2 model [13] to retrieve lifelog images efficiently in response to textual queries. Additionally, they harnessed the capabilities of the Elasticsearch engine for indexing and searching these lifelog images, utilizing the BLIP-2 embeddings to enhance the retrieval process. To streamline their data pipeline, the team incorporated chatGPT for both preprocessing and post-processing tasks. The performance of their system was evaluated, resulting in an mAP score of 0.2713, indicating the retrieval of 651 relevant items, underscoring the effectiveness of their system. Furthermore, their p@5 score achieved a value of 0.3707, confirming the system's ability to deliver valuable results, even within the first few search results, thereby highlighting its practical utility.

HCMUS-DCU LifeInsight: The LifeInsight team utilized Semantic Role Labeling (SRL)[15] to extract entities from queries and enrich metadata, harnessed Large Language Models (LLMs) to identify temporal events and create context-based prompts. They employed vector databases and vision-language pre-trained models for lifelog

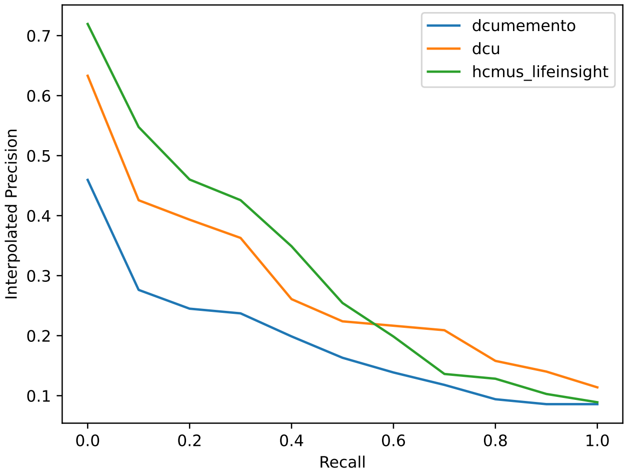

Figure 3: Comparing the best runs of the three automatic teams

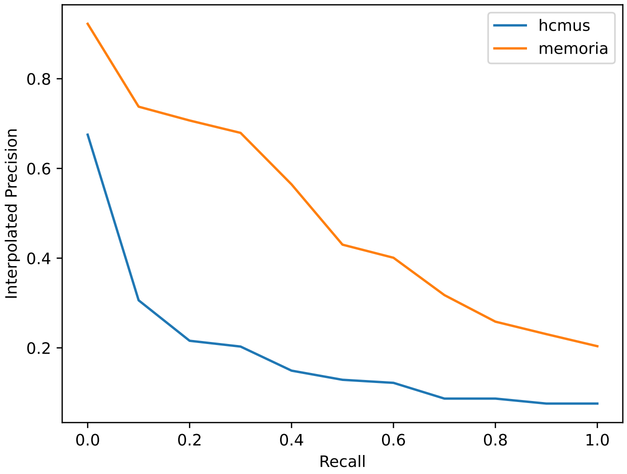

Figure 4: Comparing the best runs of the two interactive teams (HCMU-interactive & memoria-interactive)

data retrieval based on both image and text similarity. Their approach underscored the significance of subqueries in search, necessitating expertise in extracting and composing information for effective prompts, introducing global and local context-aware prompt generation for diverse query formulation, ultimately enhancing their model's contextual analysis and generative capabilities. This approach achieved an mAP score of 0.2924, retrieving 751 relevant items, and boasting a P@5 score of 0.4098, outperforming the MemoriEase and Mementosystem.

DCU-Memento system harnessed a suite of CLIP models, encompassing both OpenAI's models and the more extensive OpenCLIP models, which have been trained on a substantially larger dataset. This system utilized image-text representations derived from CLIP models to execute image search and ranking. It employed a multi-stage search process, initially ranking images based on visual descriptors and subsequently applying pertinent filters to obtain the final results. In total, the DCUMemento group submitted 9 runs, with 5 runs leveraging individual CLIP model embeddings and 4 runs applying a weighted ensemble of cosine similarity scores. The DCUMemento group, identified by the designation DCUMEMENTO_DCULSAT04_Automatic, attained an official score of 0.1734 for this task, with 450 relevant items successfully identified. The evaluation outcomes demonstrated that the ViT-g/14 model exhibited a noteworthy high recall, while the ensemble model comprising ViT-g/14 and ResNet50x64 yielded the highest overall precision. The models also demonstrated commendable performance in terms of Mean Reciprocal Rank and Normalized Discounted Cumulative Gain.

5.2 LSAT-Interactive Runs

Two groups submitted interactive runs to the LSAT sub-task, based on the outputs of two interactive retrieval systems.

HCMUS-LifeInsight system was developed to address the challenges of retrieving lifelog data and employed two distinct approaches for user interactions: Approach 1, using the BLIP model, and Approach 2, utilizing the CLIP model. The system's evaluation encompassed both expert and novice users, revealing varying levels of success in retrieving relevant images and effectively ranking them. Expert users generally outperformed novices, although some novices achieved promising results. Approach 1 resulted in a higher number of retrieved and relevant images, while Approach 2 required fewer retrieved images but achieved comparable precision. In addition to performance metrics, the system's user experience was assessed through a questionnaire, which indicated that users found the system to be supportive, efficient, exciting, and interesting. However, they stated that there is room for improvement in terms of ease of use and staying at the forefront of technology. The system submitted 8 runs, with the top-performing run being HCMS-INTERACTIVE-07, which achieved a Mean Average Precision (MAP) score of 0.1686 over 41 topics within the submission timeframe.

UA Memoria team actively participated in the competition by submitting interactive entries, wherein MEMORIA leveraged a diverse array of methodologies and models for the purpose of image annotations and retrieval. These approaches encompassed the utilization of advanced object detection and object comprehension models such as YOLOv7 and GRiT, optical character recognition (OCR) facilitated by CRAFT, scene understanding rooted in Places365, and the automated generation of descriptive captions through the employment of ClipCap. Additionally, their efforts were notably centered around event segmentation, a procedural endeavor aimed at structuring lifelog data into discernible events, thereby bolstering the organization and retrieval of memories. To accomplish this, the team adopted a hierarchical event segmentation methodology, which categorized events based on multifarious criteria, including temporal aspects (days, times of day), geographic locations, environmental conditions, and image similarity.

The team submitted a single entry that markedly outperformed other interactive submissions, as illustrated in Figure 4. This is underscored by the noteworthy achievement of achieving the highest mAP score of 0.5968 for known-item tasks and 0.2895 for ad-hoc tasks.

6 DISCUSSION AND FUTURE PLANS

The findings emerging from the NTCIR runs, as well as associated efforts within the Lifelog Search Challenge, suggest a clear trend: clip-based and Blip-based systems consistently outperform traditional concept-based approaches when dealing with multimodal lifelog archives. This observation underscores the increasing relevance and effectiveness of models grounded in the capabilities of CLIP and BLIP, underscoring their superiority in handling the complex nature of lifelog data. Moreover, the integration of the LLM model for query analysis, entities identification, and context awareness prompting has proven to be instrumental in enhancing the final ranking of results. This multifaceted approach leverages the power of advanced models for a more nuanced understanding of user queries and context, ultimately leading to improved retrieval and ranking performance.

According to the evaluation results, the HCMUS-DCU-LifeInsight and DCU-MemoriEase interactive lifelog systems have demonstrated exceptional performance. Their adept utilization of state-of-the-art models and techniques positions them as formidable contenders in the field of lifelog data retrieval and underscores their ability to meet the multifaceted challenges presented by this domain.

For interactive runs, the searcher's level of expertise consistently assumes a pivotal role in shaping the performance of these systems. An exemplar of this phenomenon can be observed through the Memoria interactive system, which has distinguished itself as the foremost performer in this task. It is worth noting, however, that the Memoria system's success is further enhanced by its judicious utilization of state-of-the-art object detection and object comprehension models, which contribute significantly to the optimization of the final ranking outcomes.

The present evaluation methodology reveals certain limitations, primarily stemming from the selection of the top 100 rankings per query as the basis for assessment. This approach may introduce inaccuracies, especially when the final ground truth for a specific query exceeds the threshold of 100 items. Consequently, the accuracy of the evaluation process becomes compromised under such circumstances. To address this issue, in the context of NTCIR lifelog6, a more adaptive and context-sensitive evaluation framework has been adopted. The number of submissions is now determined and varied in response to the specific requirements of the ground truth data, ensuring a more precise and comprehensive evaluation that aligns with the nuanced characteristics of the lifelog retrieval task.

Furthermore, we aim to investigate the prospects of specialized-domain lifelog challenges, as exemplified by the pilot NTCIR-16 RCIR task. This pilot task delved into the examination of how biometric indicators of reading comprehension influence the process of textual information retrieval. In our future endeavors, we aspire to conduct a pilot task centered around quantified self-assessment.

In addition, we intend to explore possibilities pertaining to diary search tasks and other personal data sources that possess a richer textual content. The enduring central challenge concerning interactive access to multifaceted lifelog data will continue to find its platform within the framework of the LSC workshop series. [10, 11].

7 CONCLUSION

This paper primarily focuses on providing an extensive account of the data and activities associated with the lifelog-5 LSAT sub-task as part of the NTCIR-17 event. It is important to note that, there was a dearth of research engagement with respect to the LQAT and LIT sub-tasks. While it may be challenging to derive definitive conclusions from the current findings, it is evident that a considerable volume of research remains imperative for the advancement of annotation and search tools tailored for lifelog archives. Looking ahead, our intention is to sustain and expand the lifelog task into subsequent iterations, as exemplified by the prospective NTCIR-18 Lifelog-6. In doing so, we aspire to attract greater participation in the LQAT and LIT tasks, fostering a broader research community's involvement in these critical lifelog-related endeavors. This concerted effort aims to address the pressing research needs within this domain and contribute to its ongoing development and refinement.

REFERENCES

[1] Minh-Son Dao, Duc-Tien Dang-Nguyen, Michael Riegler, and Cathal Gurrin. 2017. Smart Lifelogging: Recognizing Human Activities using PHASOR. In *International Conference on Pattern Recognition Applications and Methods*. 761–767.

[2] Martin Dodge and Rob Kitchin. 2007. 'Outlines of a world coming into existence': pervasive computing and the ethics of forgetting. *Environment and planning B: planning and design* 34, 3 (2007), 431–445.

[3] Aiden R. Doherty, Chris J.A. Moulin, and Alan F. Smeaton. 2011. Automatically assisting human memory: A SenseCam browser. *Memory* 7, 19 (2011), 785–795.

[4] A R Doherty, K Pauly-Takacs, N Caprani, C Gurrin, C J A Moulin, N E OĈonnor, and A F Smeaton. 2012. Experiences of Aiding Autobiographical Memory Using the SenseCam. *Human-Computer Interaction* 27, 1-2 (2012), 151–174.

[5] Jim Gemmell, Gordon Bell, Roger Lueder, Steven Drucker, and Curtis Wong. 2002. MyLifeBits: Fulfilling the Memex Vision. In *Proceedings of the Tenth ACM International Conference on Multimedia* (Juan-les-Pins, France) *(ACM Multimedia '02)*. ACM, New York, NY, USA, 235–238.

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

[7] Cathal Gurrin, Hideo Joho, Frank Hopfgartner, Liting Zhou, Rashmi Gupta, Rami Albatal, and Duc Tien Dang Nguyen. 2017. Overview of NTCIR-13 Lifelog-2 Task. In *Proceedings of The 13th NTCIR Conference Evaluation of Information Access Technologies* (Tokyo, Japan). National Institute of Informatics.

[8] 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 Evaluation of Information Access Technologies* (Tokyo, Japan). National Institute of Informatics.

[9] Cathal Gurrin, Björn Þór Jónsson, Duc Tien Dang Nguyen, Graham Healy, Jakub Lokoc, Liting Zhou, Luca Rossetto, Minh-Triet Tran, Wolfgang Hürst, Werner Bailer, et al. 2023. Introduction to the sixth annual lifelog search challenge, LSC'23. In *Proceedings of the 2023 ACM International Conference on Multimedia Retrieval*. 678–679.

[10] Cathal Gurrin, Björn Þór Jónsson, Klaus Schöffmann, Duc-Tien Dang-Nguyen, Jakub Lokoč, Minh-Triet Tran, Wolfgang Hürst, Luca Rossetto, and Graham Healy. 2021. Introduction to the Fourth Annual Lifelog Search Challenge, LSC'21. In *Proceedings of the 2021 International Conference on Multimedia Retrieval*. 690–691.

[11] Cathal Gurrin, Tu-Khiem Le, Van-Tu Ninh, Duc-Tien Dang-Nguyen, Björn Þór Jónsson, Jakub Lokoč, Wolfgang Hürst, Minh-Triet Tran, and Klaus Schöffmann. 2020. Introduction to the Third Annual Lifelog Search Challenge (LSC'20). In *Proceedings of the 2020 International Conference on Multimedia Retrieval*. Association for Computing Machinery, New York, NY, USA, 584–585. https://doi.org/10.1145/3372278.3388043

[12] Makoto P. Kato, Hiroaki Ohshima, Ying-Hsang Liu, and Hsin-Liang Chen. 2022. Overview of the NTCIR-16 Data Search 2 Task. In *Proceedings of the NTCIR-16 Conference*.

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

[14] Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C Lawrence Zitnick. 2014. Microsoft coco: Common objects in context. In *European conference on computer vision*. Springer, 740–755.

[15] Lluís Màrquez, Xavier Carreras, Kenneth C Litkowski, and Suzanne Stevenson. 2008. Semantic role labeling: an introduction to the special issue. , 145–159 pages.

[16] 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.

[17] Liting Zhou, Cathal Gurrin, Graham Healy, Hideo Joho, Thanh-Binh Nguyen, Rami Albatal, Frank Hopfgartner, and Duc-Tien Dang-Nguyen. 2022. Overview of the ntcir-16 lifelog-4 task. In *Proceedings of the 16th NTCIR Conference on Evaluation of Information Access Technologies*. National Institute of Informatics, 130–135.