THUIR-LL at the NTCIR-16 Lifelog-4 Task: Enhanced Interactive Lifelog Search Engine

Zhiyu He, Jiayu Li, Wenjing Wu, Min Zhang*, Yiqun Liu, Shaoping Ma
Department of Computer Science and Technology, Institute for Artificial Intelligence,
Beijing National Research Center for Information Science and Technology, Tsinghua University, Beijing 100084, China
hezy18@mails.tsingghua.edu.cn,z-m@tsinghua.edu.cn

ABSTRACT

With the development of digital information storage technology and portable sensing devices, users are gradually accustomed to recording their personal life (i.e., lifelog) in various digital ways. Therefore, the retrieval of lifelogging has become a new and essential research topic in related fields. Unlike traditional search engines, in lifelog, text and other data automatically recorded in real-time by sensors bring challenges to data arrangement and search. As the dataset is highly personalized, interactions and feedback from users should also be considered in the search engine. This paper describes our interactive approach for the NTCIR-16 Lifelog-4 Task. The task is to search relevant lifelog images from the users' daily lifelog given an event topic. A significant challenge is how to bridge the semantic gap between lifelog images and event-level topics. We propose a framework to address this problem with a multi-functional and flexible feedback mechanism and result presentation for interaction in a search engine. Besides, we propose a query text parsing procedure that parses the long query text into keywords and fills the fields automatically. We analyzed the interactive lifelog search engine with 12 topics constructed by ourselves according to LSC'18 development topics. Finally, we achieved an official result of 741 at the NTCIR-16 Lifelog-4 task in terms of RelRet score over 48 topics.

CCS CONCEPTS

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

KEYWORDS

Lifelogging, Interactive Search Engine, Information Retrieval

ACM Reference Format:
Zhiyu He, Jiayu Li, Wenjing Wu, Min Zhang, Yiqun Liu, Shaoping Ma. 2022. THUIR-LL at the NTCIR-16 Lifelog-4 Task: Enhanced Interactive Lifelog Search Engine. In *Proceedings of Proceedings of the 14th NTCIR Conference on Information Access Technologies (NTCIR '16)*. ACM, New York, NY, USA, 8 pages. https://doi.org/XXXXXXX.XXXXXXX

*Min Zhang is the corresponding author.

NTCIR '16, June 03–05, 2022, Tokyo, Japan

ACM ISBN 978-1-4503-XXXX-X/18/06…$15.00
https://doi.org/XXXXXXX.XXXXXXX

TEAMNAME

THUIR-LL

SUBTASKS

Lifelog-4

1 INTRODUCTION

Lifelogging refers to a personal multi-modal digital record obtained by various sensors and permanently stored, which can reflect the personal life experience of the recorder of the lifelogger (starting now referred to as "user") in all aspects [4]. Data such as step records, GPS positioning information, vital signs records, and photos of life scenes taken by portable cameras are all included in the scope of lifelogs. With the development of portable sensing devices such as smartphones and smart bracelets, the acquisition and storage of lifelogs have become more and more convenient, and the behavior of the lifelog recording has become more and more popular in daily life. Relevant research analysis shows that lifelog records have great potential in understanding personal life status, reflecting users' physical and mental health status, and providing practical personal advice about health [1, 8].

As a producer of lifelogging information, users sometimes need to use the lifelog data to recall details of specific scenes and understand their life status. Unlike traditional search data, lifelogs are multi-modal and unstructured data with tremendous information, including noisy and missing data, making it challenging to apply existing information retrieval methods directly. Besides, the information needs related to lifelogs are often complex. Users need to describe their needs in complex language or even multiple rounds of interaction. Also, lifelog is highly personalized, and the result is generally complex with noise. Thus, users' feedback is a critical help. Therefore, an interactive search engine with efficient interactive design and query text parsing should be highly concerned.

In recent years, many related tasks and challenges have been proposed to introduce problems in lifelog search to the research community. NTCIR-16 hosted the fourth running of the lifelog task, which aims to advance state-of-the-art research in lifelogging as an information retrieval application [3].

We participated in LSC'20 and proposed an interactive lifelog search engine. The text-based search module gives correct results on more than 60% of the development topics at LSC'20 [9]. However, to construct a text-based image search engine, we faced a vast semantic gap between multi-modal lifelog data and long query text. The contents information is much more complicated than the visual feature and objects detected from the image on the image side. On the query side, although the information needs are complex in the

lifelog search, people tend to construct their query concisely and generally to describe.

This paper proposed a framework to address this problem based on our search engine at LSC'20. New components to bridge the semantic gap are as follows:

- We optimize a query text parsing procedure. The parsed words are pressed close to the information detected from the images.
- A new feedback mechanism is proposed with ternary feedback and negative keywords in the specific numeric fields.
- We design a result presentation for interaction that can show relevant images in T-shape fixation distribution. Timeline viewing is added to provide temporal information.

Next, we analyze the interactive lifelog search engine with 12 topics constructed by ourselves according to LSC'18 development topics. It shows the significant improvement of our system both for a novice and an expert user and proves the importance of the enhancement. Our enhanced interactive lifelog search engine achieved an official result of 741 at the NTCIR-16 Lifelog-4 task in terms of RelRet score over 48 topics.

2 RELATED WORK

Image retrieval has seen a significant increase in interest over the past decade. Most traditional image retrieval methods utilize some method of adding metadata such as captioning, keywords, titles, and descriptions to the images so that retrieval can be performed over the annotation words. A content-based image retrieval (CBIR) system is required to effectively and efficiently use information from image repositories [2]. It retrieves relevant images from an image database based on primitive or semantic image features extracted automatically. Meanwhile, human perception subjectivity to show relevant feedback is incorporated into the retrieval process [12]. However, image retrieval is mainly based on large-scale image datasets, but lifelog search is built on the personal dataset, which is relatively small in size. Thus, lifelog search face a significant challenge than image retrieval. Besides, the user study on image search shows how the presentation affects the user's behavior. In Xie et al.'s work [15], instead of the traditional "Golden Triangle" phenomena in the user examination patterns of general web search, they observe a middle-position bias based on the eye-tracking study to investigate users' examination behavior in image searches. Findings from user studies can inspire the design of the lifelog search engine. Therefore, the lifelog search engine should pay more attention to user interaction and understanding of information needs.

With an increasing number of workshops and tasks on lifelogging, there is already some research about the lifelog search engines in recent years. NTCIR-14 Lifelog-3 held Lifelog Semantic Access sub-Task (LSAT); we know this is a recent similar task at NTCIR. Some of the systems are automatic with the help of external data on the web and the preprocessing of the lifelog dataset [13]. More designed to be an interactive system [11]. The best performing run came from HCMUS [6], which used a custom annotation process for the lifelog data based on the identifiable habits of the lifeloggers. Lifelog Search Challenge (LSC) focuses on the interactive system. The Myscéal retrieval system [14] was the top-performing system developed at LSC'21. It explored query expansion and word embedding approaches to interactive retrieval and enhanced modules like map position and day summary for the novice. When considering the feedback mechanism, the Exquisitor system [5] proposed a novel way of simply using relevant feedback from the user to find results. It uses binary feedback for each item and trains the classifier to provide a new round of recommendations. Besides, LifeXplore [7] combined chronologic day summary browsing with interactive combinable concept filtering. These modules bring new possibilities for interactive systems.

In our work, we propose a multi-modal interactive search engine. We focus on the query text parsing procedure and the interactive manners to bridge the semantic gap.

3 LIFELOG DATA AND FEATURE EXTRACTION

The NTCIR-16 lifelog-4 organizers reused an existing dataset from LSC'21, a multi-modal dataset from one active lifelogger. It contains 114-day multi-modal lifelog data captured and synchronized from both smartphones and multiple sensors recorded continuously in 2015,2016,2018 [3]. The lifelog images are fully anonymized to prevent any personal information from leaking. The organizers provided the metadata enriched to provide descriptive and temporal information for each moment. Besides, the visual concept extracted from the non-redacted version of the images by the model pretrained on the COCO dataset.

Following the previous work [9] at LSC'20, we cluster images according to histogram similarity. As photos in the same group are taken in a similar scene, each group is called a **shot**. At last, we have 50233 shots of denoised images. A shot is treated as an atom unit for feature generation and search, which substantially reduces our search engine's computation. After that, we conduct multilevel feature extraction [9] to get visual features, textual features, and behavior features.

4 INTERACTIVE SEARCH ENGINE

In this section, we introduce our interactive search engine. Section 4.1 proposes the framework of our lifelog search engine with the search mechanism. After the overview, we highlight the essential components of bridging the semantic gap of our content-based search engine. In order to retrieve the target, the user may generally modify the query in the lifelog process. So it is necessary to provide a way for the text parser. Section 4.2 summarizes the procedure of query text parsing. Besides, as an interactive search engine, we pay great attention to the functionality and convenience of the user's feedback mechanism, which is described in Section 4.3. In addition, a previous user study shows the impact of presentation on efficiency. So we design an interactive presentation based on lifelog scenario as shown in Section 4.4.

4.1 Overview of the system

The framework of our system is shown in Figure 1.

Section 3 introduces the datasets and feature generation methods. Each shot is represented as a document with abundant content features from visual, textual, and behavior in our system. Each feature corresponds to a facet in the search engine schema.

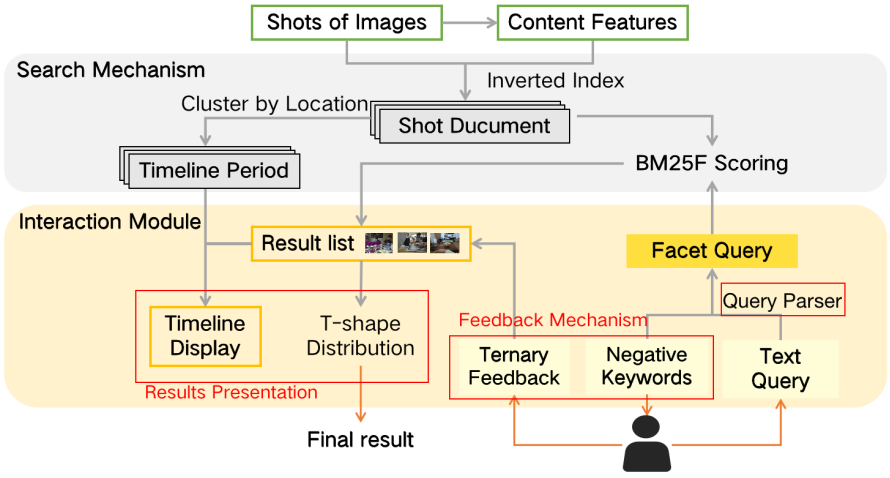

Figure 1: The framework of our lifelog search system. The enhancement points are query parsing procedure, feedback mechanism, and result presentation.

We use Whoosh[1] to build our search modules, and the inverted index is employed to save documents. Besides, for each day in lifelog, neighboring shots with the same location are clustered as a time period, and objects with the highest confidence score are saved for each timeline period.

We also build a dictionary of time, objects, location, and activity features in our dataset. As the user give a textual query, it is first parsed into facets and looked up in the feature dictionary. The category-directed query parser is described in detail in Section 4.2. All the facets are combined with logical AND, while logical OR is used to connect multiple query keywords within locations, activities, and objects. We Implement the BM25F scoring algorithm by Whoosh. Based on the BM25 algorithm, we calculate and sum each word separately in each field. With the faceted query, the BM25F scoring function is then used to search for Top N relevant shots, and results are listed in T-shape fixation distribution by their relevance scores.

As an interactive search engine, the user can check the shots and even view the timeline of the day of a particular shot. After the checking, the user should give ternary feedback (exclude, choose, or keep) of every result shot. The user can further modify the query and input negative keywords in specific numeric fields, which commonly update the facet query. Then, the search engine will present a new result list based on the facet query by the BM25F scoring function. The flow path may continue until the user finds the relevant image. The images chosen by the user are the final result.

Besides feature generation, the process of building shot documents and timeline period is also the same as our lifelog search engine in LSC'20 [9]. The BM25 scoring function by the faceted query to get relevant shots also does not change.

Our improvements are mainly in the interaction module, which is summarized into query text parsing, feedback mechanism, and result presentation for interaction, and are detailed in the following three subsections, respectively.

4.2 Query text parsing

Since the user's input is a piece of free text, the index document is stored according to the keyword field. In order to improve search efficiency and accuracy, it is necessary to extract the practical information of each field from the free text to form a query. Therefore, a text query parser is designed. The tense and syntactic structure are concerned. Besides, rules are added to generate corresponded noun phases to press close to the objects detected from the images.

After the free text segment is input into the parser, it first needs to be lowercase, stemmed, and filtered out of stop words to generate normalized sentences. The normalized statements are processed through clauses and become a sequence of statements. For each sentence in the sequence, the parser does the following procedure.

First, we determine the tense of the sentence according to the verb tense. The tense will affect the field matching of object labels, locations, and activities. Passive voice needs to be treated separately since it is expressed in the form of the "verb be + past participle".

Then, certain words are filled according to the tense. If the sentence contains time information and vital sign data information, we extract and fill in the corresponding fields of the query. Moreover, we loaded the dictionary stored when constructing the feature index because relatively few locations and activity features are available. Thus, we matched the locations and activities one by one in the statement and combined the locations and activities into the corresponding fields that match the tense.

Finally, the parser needs to parse the object tags. As shown in Table 1, we generate a syntactic tree and classify the object label descriptions in the user text input into three categories according to the test callout. These three types of phrases need to be matched in the sentences, respectively, and the successfully matched phrases extract their **JN** type noun phrases and fill them into the object label field that matches the tense in the query. Besides, rule-based

[1]https://pypi.org/project/Whoosh/

noun-category correspondence is used to extract the category of the nouns. For example, we extracted "people" when there are people-related words such as "woman" in the text and "food" when eating-related words such as "sushi" in the text. It is a crucial step to match the natural language to detected objects in images.

After the above steps, the parser can extract the most valid information from the user input text. Two examples are shown in Table 2.

4.3 Feedback mechanism

This subsection discusses the feedback mechanism we provide to the users.

In a round of interaction, it is undoubtedly a more efficient way for users to present more information with less behavior. There are generally three types of user feedback presented to a particular shot: **exclude** (i.e., exclude the irrelevant shot), **choose** (i.e., choose the relevant shot), and **keep** (i.e., keep the shot to the next round if the user cannot immediately judge whether it is relevant). Ternary feedback can be seen in Figure 2. Compared with binary feedback, ternary feedback provides users with more fault tolerance, which is novice-friendly. Meanwhile, we provide one of three types of feedback as the default item. Considering that the actual relevant results in the lifelog search scenario are presented with much noise, we use **exclude** as the default item.

After adding ternary feedback, the user quickly browses through all shots and chooses **choose** or **keep** for potentially relevant results, making the user participation in the system more efficient.

In the face of the semantic gap, it is not enough to only provide feedback for each result. The low-level information brought by images will be inconsistent with human cognition. Thus the system may recall many wrong results with wrong keywords. Thus, as shown in Figure 2, negative feedback keyword boxes are on the left side of the result page. Users can modify the query and add negative feedback keywords in a specific field. This step removes noise significantly.

4.4 Result Presentation for Interaction

The presentation of results and the information offered to the users affect how users find relevant results on search engine result pages [10].

Ignored the content of image results (e.g., visual saliency), the dominant position of first arrival time and examination duration exists in the examination process. Presenting more relevant results in these locations impacts user experience and efficiency. Previous work[15] conducted an eye-tracking study to investigate users' examination behavior in image searches and observed a middle-position bias in the horizontal direction instead of the traditional "Golden Triangle" phenomena in the user examination patterns of general Web search. Based on their conclusions and recommendations, we designed the "T-shape" fixation distribution instead of the "F-shape".

Figure 2 shows the interface of the result page. The numbers in orange were marked to present the hidden relevance ordering, which demonstrates the correspondence between ordering and position. In the horizontal direction, the shots diverge from the middle to the sides. Consistent with the standard browsing order, the images are arranged from top to bottom in a vertical direction. The above is called a T-shaped distribution. Since the user's attention tends to be in the middle of the horizontal line, we put the most relevant image here so that the user will focus on it.

In addition to the distribution of the results, we also pay attention to the user's understanding of the results. In the lifelog scenario, each picture does not exist independently. Instead, a series of pictures presents a complete behavioral story. Thus, we add the timeline to give the user a more straightforward way to understand the background of a single shot in general. The user can click on the 'view' button at specific shots. The button can be seen in Figure 2. Then the location tracking, dwell time, and relevant locations during that day will be displayed, shown in Figure 3. The user can further click on any location to view key images of the shots at the location in the corresponding time period. The timeline for the day of a particular photo implies context information that the image cannot present at a single moment. The timelines provide more features to pictures, bridging gaps in interactive systems. Based on the timeline for the day of a particular photo, the user can analyze whether the photo is relevant or not. Timeline viewing is an essential part of making the system convenient and novice-friendly.

5 EXPERIMENTS AND RESULTS

This section analyzes the key components of the lifelog retrieval system. Specifically, we compare the search engine before and after the interactive methods enhancement based on the constructed topics. An expert and a novice user conduct the experiments as the people involved in the system.

5.1 Topic Construction

Following the topics in NTCIR-16 Lifelog-4, we construct 12 topics in traditional TREC style, with title, description, and narrative, to verify our system. Our system is enhanced based on our LSC'20 system [9]. Previous work used the LSC'18 topics for user experiments. In order to better compare with the previous system, all the topics we constructed this time refer to the LSC'18 topics.

Overall, there are six KNOWNITEM topics and six ADHOC topics. KNOWNITEM is the topic with relevant images from one or few moments. We choose from the LSC'18 development challenge [2] with the description is as briefly as possible. The selection criteria are that even in the three-year dataset, the topic pointed to a uniquely relevant moment and that relevant images for a specific topic could be easily found based on the exact time. However, the topic does not show the accurate time in the actual situation. Thus, finding the relevant result can be challenging without knowing the time. Finally, topics numbered 02, 04, 14, 15, 20, and 24 in LSC'18 were chosen. As for ADHOC, the topic has some relevant images from various moments. We were inspired by LSC'18 development topics numbered 01, 03, 06, 08, 11, and 12. These new themes are built with the same actions in different scenarios. It's not enough to find relevant images on a specific date.

The differences between KNOWNITEM and ADHOC topics can not be described as the differences in the number of their images.

[2] http://lsc.dcu.ie/2018/resources/LSC2018_dev_topics.xml

Type	Grammar	Explanation	Example
JN	<DT>*(<CC>*<JJ>*)*<NN>	noun phrases containing adjectives	a white and blue shirt
PN	<JN><IN>+<JN>	noun phrases joined by prepositions	a white cat on a chair
NJ	<JN><VB><IN>*<JN>+	noun phrases joined by verbs	a man sitting on a chair

Table 1: Three types of text description methods for object labels.

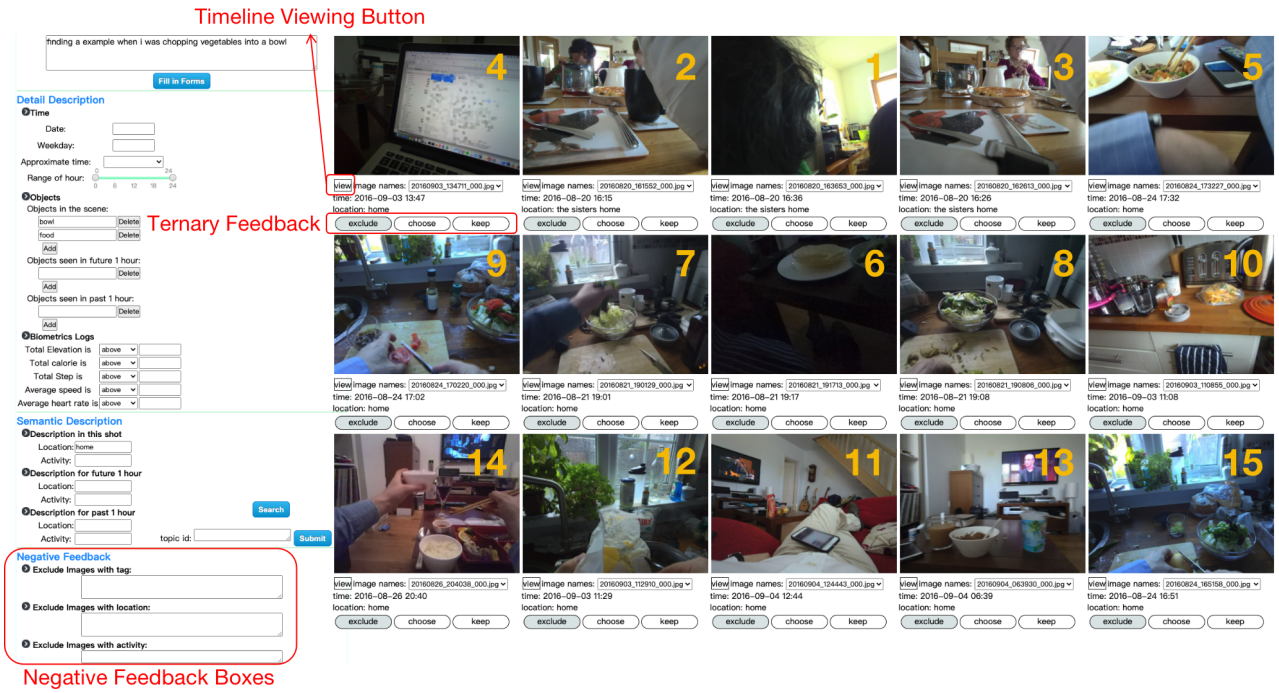

Figure 2: The Result Page of the System. The returned results are in the "T-shape" distribution. Numbers in orange were marked to present the relevance order. Ternary feedback buttons, negative feedback boxes, and timeline viewing buttons are marked red.

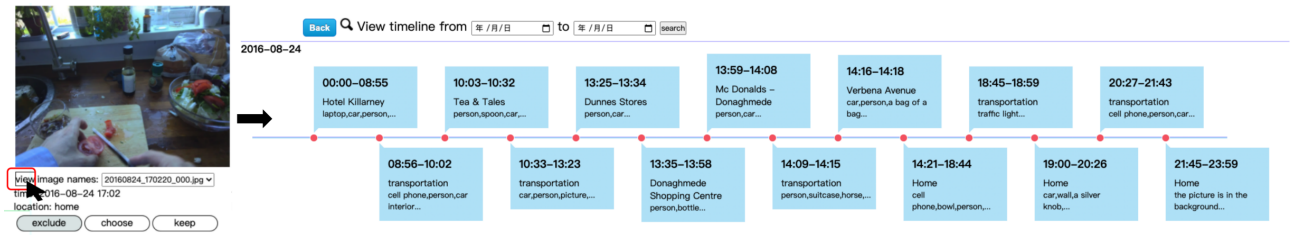

Figure 3: Interface for Timeline viewing. When the user clicks the "view" button of the target shot on the result page, it will jump to the timeline of that day.

They are different mainly in the topic frequency. Though the moment of a KNOWNITEM topic only happens once a year, it may last a few minutes, so the total number of relevant images can be more than the ADHOC topics.

Example topics are shown in Table 3 and all the topics constructed are displayed with their titles in Table 4.

5.2 Experiment Settings

Based on the constructed topics, We compare the search engine before and after the improvement by both a novice and an expert user. The novice is a recruited college student who is a proficient web search user. She had no previous exposure to the lifelog and was not involved in the design of our search engine. The novice user explored the operation of the search engine for several days

Query Text	Facet Query
Find examples of when I was lifelogged when eating lunch at work in my office.	Tags in this shot: [food, office] Approximate time: [noon] Location in this shot: [office, work]
Find examples of when I was in meetings with other people in rooms with red carpet. Before that, i had lived in hotel.	tags in this shot: [red carpet, room, rooms, people] tags in past 1 hour: [hotel] location in past 1 hour: [hotel]

Table 2: The examples of the parsed words by the text parser.

Type	KNOWNITEM	ADHOC
Title	Building a Chair	Chopping vegetables
Description	Find the example when I was building a chair that is wooden in the late afternoon.	Find the example when I was chopping vegetables into a bowl.
Narrative	To be relevant, the images must show the lifelogger building a chair at work, in an office environment. It happened in the late afternoon.	To be relevant, the images must show the lifelogger chopping vegetables. Bowls can be seen in the images.

Table 3: The detailed examples of KNOWNITEM and ADHOC show the differences between them: The KNOWNITEM topic only happen once though it has a lot of relevant images, while the ADHOC Topic is shown at different times in the dataset.

Titles of All Topics	
KNOWNITEM	**ADHOC**
Building a Chair	Coffee in Helix
Scanning Receipts	Watching Foodball
Suitcase in car	Photos of white building
Stone Castle	Chopping vegetables
Sunrise Photo	Waiting for the Train
Hair Salon	Tidying Garden

Table 4: The above shows the titles of all the constructed KNOWNITEM topics and ADHOC topics.

after the developer explained the background and search process to her. During the training process, the novice only saw some lifelog images that were not in the search target and did not know the entire content of the dataset. The expert user comes from the development team and is also a college student. It has to be admitted that in the process of development, experts inevitably learn about the characteristics of the data.

Before the experiment, both the novice and the expert familiarize themselves with the system through other topics. Then, each user tests both systems for the same topic. Considering the distraction of changes in user familiarity with the same topic in the experiment, we provide an example image for each topic. The example is a typical relevant image, which is one of the images that the user is looking for. We only show the example image without textual features such as date. The users could not retrieve images through the visual features of the example. Tests were performed consecutively for each topic on the pre-enhanced and post-enhanced systems. The order of the two systems is random.

The user pastes **description** of the topic in the search box and clicks the "fill in form" button. Then the timing begins. The user can modify the parsed query and do each round of interaction. When each round of interaction ends, the time of the timer is the **elapsed time** of the chosen shots on the current page. The user needs to view each picture in the chosen shot during the process and mark the ImageIDs that are not adopted in the shot. After the experiment is completed, we record the ShotID corresponding to all ImageIDs. A maximum of 100 images can be returned per topic.

We use the following evaluation metrics to measure the system performance:

- **num_q** represents the number of the returned topics.
- **num_ret** represents the number of returned documents.
- **num_rel_ret** means the number of correctly predicted relevant results among the returned documents.
- **map** is the mean average precision, which is the arithmetic mean between topics. If the set of relevant documents for a topic t_j is $d_1, \ldots, d_{m_j}$ and R_{jk} is the set of documents retrieved until the user gets d_k, then:

$$MAP(T) = \frac{1}{|T|}\sum_{j=1}^{n}[\frac{1}{m_j}\sum_{k=1}^{m_j} Precision(R_{jk})]$$

- **gm_map** is the geometric mean:

$$gm\_MAP(T) = \sqrt[n]{\prod_{j=1}^{n}[\frac{1}{m_j}\sum_{k=1}^{m_j} Precision(R_{jk})]}$$

- **Rprec** is the mean of the precision after R documents have been retrieved, where R is the number of relevant documents for the topic.
- **recip_rank** represents the mean of the reciprocal of the position of the first related document for each topic.
- **P_k** is the proportion of correctly predicted relevant results from all returned results in the top k documents.

Specifically, as for the KNOWNITEM topics, we considered the contents in the **RelevantImageIDs** label as relevant images. There are 53 relevant images from 6 KNOWNITEM topics. As for the ADHOC topics, relevant images are from validation screening of all images found by the expert and reaching a consensus with the novice. Finally, there are 178 relevant images from 6 AdHOC topics in total. After sorting the results by **elapsed time**, the **similarity** of each result is sorted backward from 100. Then we used trec_eval to evaluate, following NTCIR-16 Lifelog-4.

5.3 Overall Result

The overall results are shown in Table 5.

Intuitively, the post-enhanced system improves most metrics than the pre-enhanced system for both the expert and the novice.

	KNOWNITEM				ADHOC			
	Novice		Expert		Novice		Expert	
	pre	post	pre	post	pre	post	pre	post
num_q	6	6	6	6	5	5	6	6
num_ret	24	26	37	40	215	233	159	162
num_rel_ret	19	21	31	36	51	66	118	103
map	0.2086	0.3753	0.5711	0.6410	0.3047	0.4857	0.4850	0.7160
Rprec	0.2086	0.3753	0.5842	0.6410	0.2556	0.5220	0.5487	0.7571
recip rank	0.6667	0.8333	1.0000	1.0000	0.6848	0.6917	0.5931	0.8431
P_5	0.3333	0.4000	0.5667	0.6000	0.5600	0.6400	0.5000	0.8000
P_20	0.1583	0.1750	0.2583	0.3000	0.4300	0.4300	0.4917	0.6000
P_100	0.0317	0.0350	0.0517	0.0600	0.1020	0.1320	0.1967	0.1717

Table 5: The results of evaluation metrics to measure the pre-enhanced and the post-enhanced system by a novice and an expert user for KNOWNITEM and ADHOC topics.

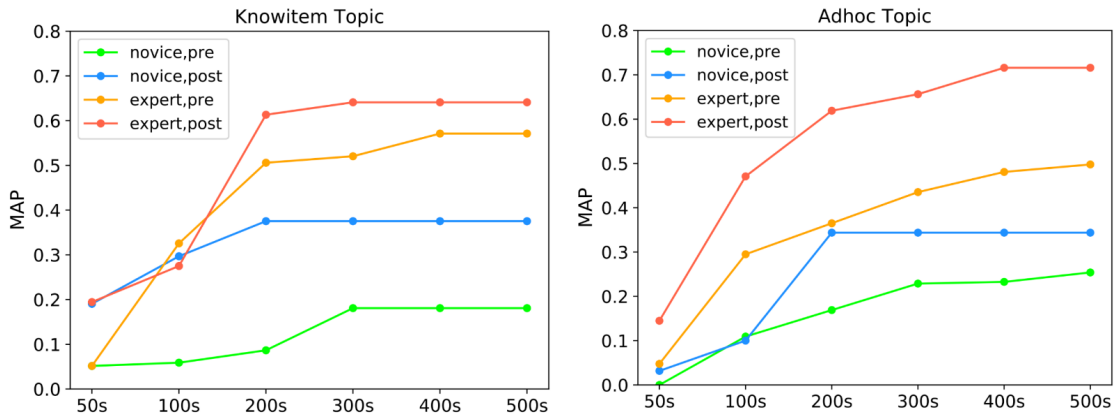

Figure 4: The MAP at various time cut-offs for both KNOWNITEM topics and ADHOC topics. Each line represents the novice and the expert user test with pre-enhanced and post-enhanced system, respectively.

It illustrates the effect of enhancement on the lifelog search engine. The query parser automatically turns text query into facet query, and the feedback mechanism provides the user with an efficient and convenient way to get involved in the system. The result presentation considered user browsing habits and the need for timeline information.

However, the ADHOC topic for the expert at **num_rel_ret** is a particular case. When the expert test the ADHOC topics, two topics sensitive to user familiarity are randomly tested with the post-enhanced system first and then with the pre-enhanced system. For this reason, the number of relevant images among return images of the pre-enhanced system is more than the post-enhanced. Nonetheless, **Rprec** is boosted as it considers the average recall of each topic. **MAP**, **MRR**, and **P@k**, these position-related metrics have also improved. These results indicate that the post-enhanced system is better than the pre-enhanced one in terms of the order in which the relevant image is found, although the number of total relevant results does not necessarily improve.

Figure 4 shows the **MAP** at various time cut-offs for both KNOWNITEM topics and ADHOC topics. We evaluate the result within 100s, 200s,etc. Considering that the number of topics on which related documents are found is different in different times. **MAP** calculated by trec_eval is the mean of the mean average precision of the finding topics. To balance the difference in the number of topics, we correct the **MAP** obtained by trec_eval:

$$MAP' = MAP \times num\_q \div 6$$

In the figure 4 and the following, **MAP** refers to the revised MAP'. When considering the retrieval time, we find that the post-enhanced

system may not be better than pre-enhanced in the early stage. But the post one is better than the pre one in the later stage (>200s).

Meanwhile, we also find that the effect of the expert is significantly improved than that of the novice, which indicates that the user's familiarity with the system has a significant influence on the operation effect of the interactive system. The difference comes from whether the user can use search engines proficiently and whether the user has experience with lifelogs. Thus, this also places requirements on our search engine. A user-friendly interface should provide search convenience for experts. Also, it does not ignore the acceptance of new systems by novelties.

As for the difference between the two topic types, far more images are returned in the ADHOC topic than the KNOWNITEM, which is in line with the considerations when constructing topics. At the same time, on the topics of ADHOC, the interference caused by differences in user judgment is more apparent, and the noise caused by the test sequence of the two systems will also be tremendous. We can see that the performance of the novice on the post-enhanced system may be better than the expert on the pre-enhanced system when doing ADHOC topics. The finding shows the apparent improvement over the systems. It also gives a sight that the novice has the potential to catch up to the expert on topics with uncertain outcomes and noise. This encouraging conclusion also illustrates the irreplaceability of novice in the evaluation.

5.4 Online Performance

We test our search engine with 48 development topics at NTCIR-16 Lifelog-4. Our search engine is built on the four-month dataset released in NTCIR-16 Lifelog-4. We input **description** as the textual query for every topic. Then the system parses the query into facets, and the user searches with the feedback mechanism and result presentation described in Section 4.3 and Section 4.4. The system fails to detect topics 41, 47, and 48. It detects 1298 images in 300 seconds from other 45 topics, of which 741 are consistent with the official (there are 2986 relevant images in total).

The NTCIR organizer gives the online evaluation results.

6 CONCLUSION AND FUTURE WORK

This paper presents our multilevel lifelog search engine with novelty and reasonable interactive methods from query text parser, feedback mechanism, and result presentation for interaction.

Based on the interactive search engine where users can find images with query text in lifelog scenario [9], we proposed a framework of enhanced interactive methods to bridge the gap between lifelog images and event-level topics. Specifically, we raise the query text parsing procedure. Besides, the feedback mechanism with ternary feedback and negative keywords in a specific numeric field was used. Moreover, the result presentation for interaction is enhanced by the "T-shape" distribution of relevant images and timeline viewing function.

Our experiment on the constructed topics shows the promising progress of our enhancement for both novice and expert, which verifies the effectiveness of our query text parser, interactive mechanism, and presentation. Meanwhile, the online result of [ranking] indicates the usefulness and precision of our search engine.

In the future, optimizing methods to eliminate noise in the dataset can be a direction. More tools from computer vision and natural language processing communities will be used for object detection and semantic information extraction. The user behavior of image search engines has given us a lot of inspiration. We hope to use eye-tracking to explore how the user's attention distribution is in lifelog interactive search scenarios. A better understanding of users' search strategies and interactive behavior patterns can optimize the system.

ACKNOWLEDGMENTS

This work is supported by the Natural Science Foundation of China (Grant No. U21B2026) and Tsinghua University Guoqiang Research Institute.

REFERENCES

[1] 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. https://doi.org/10.1068/b32041t

[2] V.N. Gudivada and V.V. Raghavan. 1995. Content based image retrieval systems. *Computer* 28, 9 (1995), 18–22. https://doi.org/10.1109/2.410145

[3] Cathal Gurrin, Frank Hopfgartner, Duc-Tien Dang-Nguyen, Thanh-Binh Nguyen, Graham Healy, Rami Albatal, and Liting Zhou. 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.

[4] Cathal Gurrin, Alan F. Smeaton, and Aiden R. Doherty. 2014. LifeLogging: Personal Big Data. *Found. Trends Inf. Retr.* 8 (2014), 1–125.

[5] Omar Shahbaz Khan, Aaron Duane, Björn Þór Jónsson, Jan Zahálka, Stevan Rudinac, and Marcel Worring. 2021. Exquisitor at the Lifelog Search Challenge 2021: Relationships Between Semantic Classifiers. In *Proceedings of the 4th Annual on Lifelog Search Challenge* (Taipei, Taiwan) *(LSC '21)*. Association for Computing Machinery, New York, NY, USA, 3–6. https://doi.org/10.1145/3463948.3469255

[6] Nguyen-Khang Le, Dieu-Hien Nguyen, Trung-Hieu Hoang, Thanh-An Nguyen, Thanh-Dat Truong, Duy-Tung Dinh, Quoc-An Luong, Viet-Khoa Vo-Ho, Vinh-Tiep Nguyen, and Minh-Triet Tran. 2019. HCMUS at the NTCIR-14 Lifelog-3 Task. In *Proceedings of the 14th NTCIR Conference on Evaluation of Information Access Technologies*. 48–60.

[7] Andreas Leibetseder and Klaus Schoeffmann. 2021. LifeXplore at the Lifelog Search Challenge 2021 *(LSC '21)*. Association for Computing Machinery, New York, NY, USA, 23–28. https://doi.org/10.1145/3463948.3469060

[8] Jiayu Li, Weizhi Ma, Min Zhang, Pengyu Wang, Yiqun Liu, and Shaoping Ma. 2021. Know Yourself: Physical and Psychological Self-awareness with Lifelog. *Frontiers in Digital Health* (2021), 96.

[9] Jiayu Li, Min Zhang, Weizhi Ma, Yiqun Liu, and Shaoping Ma. 2020. *A Multi-Level Interactive Lifelog Search Engine with User Feedback*. Association for Computing Machinery, New York, NY, USA, 29–35. https://doi.org/10.1145/3379172.3391720

[10] Zeyang Liu, Yiqun Liu, Ke Zhou, Min Zhang, and Shaoping Ma. 2015. Influence of Vertical Result in Web Search Examination. In *Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval* (Santiago, Chile) *(SIGIR '15)*. Association for Computing Machinery, New York, NY, USA, 193–202. https://doi.org/10.1145/2766462.2767714

[11] Van-Tu Ninh, Tu-Khiem Le, Liting Zhou, Graham Healy, Kaushik Venkataraman, Minh-Triet Tran, Duc-Tien Dang-Nguyen, S Smith, and Cathal Gurrin. 2019. A baseline interactive retrieval engine for the NTICR-14 Lifelog-3 semantic access task. In *The Fourteenth NTCIR Conference (NTCIR-14)*.

[12] Yong Rui, T.S. Huang, M. Ortega, and S. Mehrotra. 1998. Relevance feedback: a power tool for interactive content-based image retrieval. *IEEE Transactions on Circuits and Systems for Video Technology* 8, 5 (1998), 644–655. https://doi.org/10.1109/76.718510

[13] Tokinori Suzuki and Daisuke Ikeda. 2019. Smart lifelog retrieval system with habit-based concepts and moment visualization. *Proceedings of NTCIR-14, Tokyo, Japan* (2019).

[14] Ly-Duyen Tran, Manh-Duy Nguyen, Nguyen Thanh Binh, Hyowon Lee, and Cathal Gurrin. 2021. Myscéal 2.0: A Revised Experimental Interactive Lifelog Retrieval System for LSC'21. In *Proceedings of the 4th Annual on Lifelog Search Challenge* (Taipei, Taiwan) *(LSC '21)*. Association for Computing Machinery, New York, NY, USA, 11–16. https://doi.org/10.1145/3463948.3469064

[15] Xiaohui Xie, Yiqun Liu, Xiaochuan Wang, Meng Wang, Zhijing Wu, Yingying Wu, Min Zhang, and Shaoping Ma. 2017. Investigating Examination Behavior of Image Search Users. 275–284. https://doi.org/10.1145/3077136.3080799