# A Baseline Interactive Retrieval Engine for the NTCIR-14 Lifelog-3 Semantic Access Task

Van-Tu Ninh[1⋆], Tu-Khiem Le[1⋆], Liting Zhou[1], Graham Healy[1], Kaushik Venkataraman[1], Minh-Triet Tran[2], Duc-Tien Dang-Nguyen[3], Sinead Smyth[1], and Cathal Gurrin[1]

[1] Dublin City University, Ireland
[2] VNU-HCM, University of Science, Vietnam
[3] University of Bergen, Norway

**Abstract.** This paper describes the work of DCU research team in collaboration with University of Science, Vietnam, and University of Bergen, Norway at the Lifelog task of NTCIR-14. In this paper, we describe a new baseline interactive retrieval engine based on faceted retrieval and the present the results of an initial experiment with four users. The interactive retrieval system we describe utilises the wide range of lifelog metadata provided by the task organisers to develop an extensive faceted retrieval system.

**Keywords:** Interactive lifelog search engine · Information retrieval
**Task Name:** LifeLog-3
**Subtasks:** Lifelog Semantic Access Task (LSAT)

## 1 Introduction

Information Retrieval has a long history of utilising the human as a key component of a retrieval system. Our current generation of WWW search engines rely on the human as an integral part of the search process, in terms of query generation, refinement and result selection. Inspired by the 'human-in-the-loop' model of information retrieval, the DCU team, with the support of VNU-HCM, University of Science and the University of Bergen, developed a prototype baseline interactive retrieval system for the LSAT - Lifelog Semantic Access subtask of the NTCIR-14 Lifelog task [2]. In this paper we introduce this baseline retrieval engine, we present the performance of the retrieval engine in the LSAT task, we report on the findings of a small-scale qualitative user study of the prototype, and we highlight the enhancements required for the prototype.

## 2 Related Interactive Lifelog Retrieval Systems

The Lifelog Semantic Access Task, which began in NTCIR-12, allows both automatic and interactive lifelog search systems [3] to be comparatively evaluated in

⋆ The two authors contributed equally to this paper.

an open benchmarking exercise. In NTCIR-12, the team from the University of Barcelona and Technical University of Catalonia was the only one who developed interactive search engine [10]. Their approach was to utilize the visual semantic concepts from image and use them as tags for image retrieval system. They also use WordNet to create the similarity between tags to suggest novice/expert users choose the most relevant appropriate tags. Moreover, a heatmap was generated to show the confidence of the retrieval result which aims to achieve the best configuration of precision and recall of their retrieval system. In the official results of the lifelog task, their best run (unsurprisingly) outperforms all the best ones of other teams that built automatic search engines [3].

More recently, we note the introduction of a new challenge, specifically aimed at comparing approaches to interactive retrieval from lifelog archives. The Lifelog Search Challenge (LSC) utilises a similar dataset [4] to the one used for the NTCIR14-Lifelog task. The LSC has attracted significant interest from participants and we report on the most relevant of these here. DCU's research team developed a novel approach for lifelog exploration using a virtual reality access mechanism, with different methods of user interaction to increase the speed of searching for relevant moments: distanced-based and contact-based interaction [1]. A key feature of this system was the limited use of facets for the query generation, where these facets were visual concept and time based. Another approach was taken by two participants [9] [8], both of whom converted the lifelog data into video sequences and developed retrieval systems as if the lifelog data was video sequences. Both of these systems relied heavily on the visual concepts to support retrieval.

## 3 An Interactive Lifelog Retrieval System

For the LSAT sub-task, we developed a retrieval system to provide timely, precise and convenient access to a lifelog data archive. The system, as well as our official submissions were designed to maximize recall, in order to support a user to access their life experiences in a real-world lifelogging scenario.

### 3.1 Data Preprocessing

After analyzing NTCIR-14 lifelog data [2], we divide the data into five main categories: time, location, activities, biometrics and visual concepts.

1. **Time**: For time data, we split the minute based lifelog data into selection of range of hours/minutes/days/day of week for lifelog search engine. Novice/expert user can utilise this information to narrow the scope of searching for a topic in lifelogger's data. All time data in our lifelog are converted into the UTC time standard.
2. **Location**: For location data, we also utilize timezone information to know the region/country where the lifelogger is visiting. We convert locations into semantic names to help novice/expert user locate the category of place when searching for lifelogger's moments.

3. **Activity**: The activity data contains two categories: walking and transport.
4. **Biometrics**: The biometrics data that we use in our search engine includes: heart rate and calories.
5. **Visual concepts**: We included the visual concepts that were distributed with the dataset. Visual concepts are of three types: place attribute, place category, and visual objects. The place features were extracted using places365-CNN [12]. The visual objects' categories originate from MSCOCO dataset [7] and are automatically detected using object detection network [11].

### 3.2 The Baseline System

The baseline interactive retrieval engine implemented a faceted search system in which a user could either enter a textual query in a conventional text box, or select from a range of facets of the metadata to locate items of interest. The faceted search system operated over a range of metadata which are listed in subsection 3.1 which are day of the week, date, time range, user activity (walking/transporting), biometrics data ranges (calories and heart rate), location (location category and name), and visual concepts (place attributes, place categories, and detected objects) in the corresponding order.

When searching using the conventional text box, a user is limited to utilising only visual concepts, activity, and location. If user desires to utilize all the metadata in searching for relevant items, using the facets to query can support this.

The interface, showing the faceted panel (left), the querybox (top) and the result browsing panel (right) is shown in figure 1. Note the timer on the top of the main panel, which was added to support the LSAT interactive experiment.

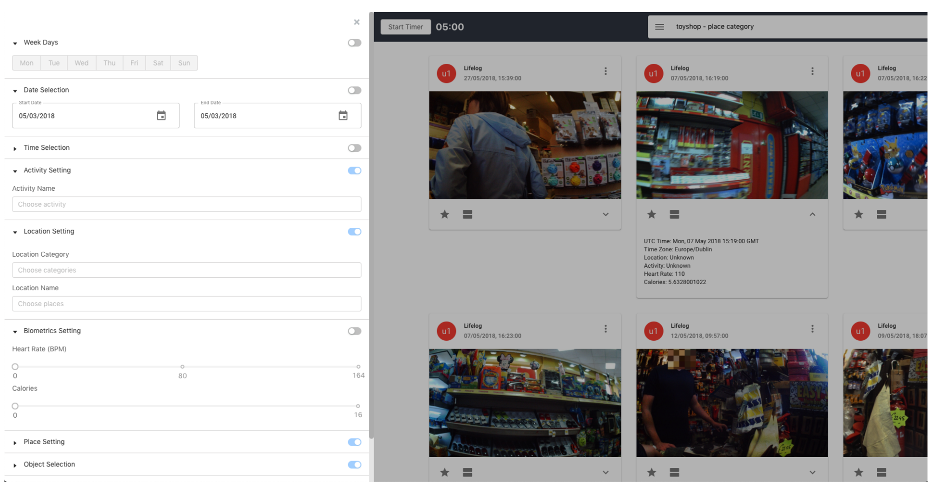


**Fig. 1.** Query Panel

Upon generating a query, the system generated a list of results (20 per page and 5 pages of results) ranked in temporal order, as shown in 2, using a conventional text ranking algorithm. The unit of retrieval was the image, as was expected for the LSAT task. Each image is given a title, date, a button to choose the image and another one to show before & after moments of the current image. Summary metadata from each image could be displayed by selecting the image. If an image was selected as being relevant (the star icon), then it was saved for submission. Submission occurred automatically after a given time period had elapsed, in our case, this was five minutes.

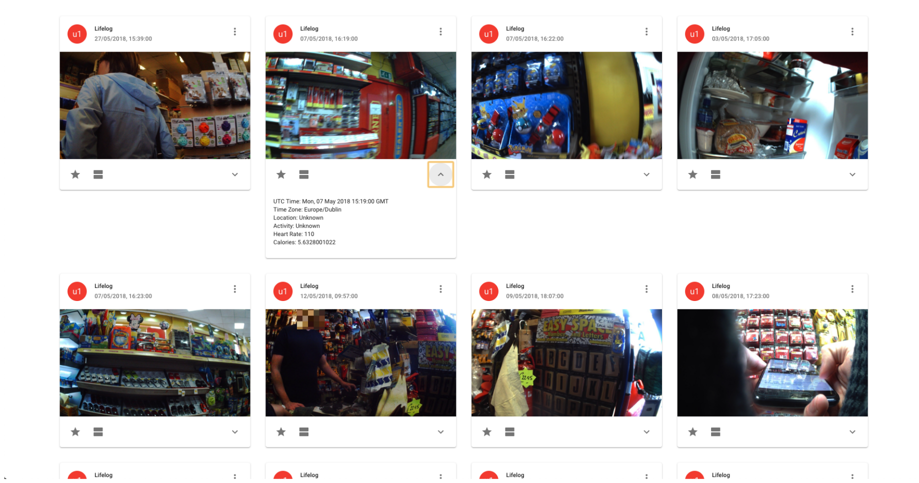

**Fig. 2.** Result Panel

Additionally, for any image, the temporal context was made available by selecting the double box under each image. The temporal context appeared as a hovering panel and the user could browse back (left) or forward (right) in time, see figure 3 for the temporal context of an image for the topic 'toystore'. Selecting an image allowed it to be flagged as relevant.

At the end of a five minute period, all saved images were used to form the official submission. Additionally, all images before and after (to a depth of ten) were appended to the end of the official submission for evaluation. The idea was that additional relevant content could be found in the temporal neighbourhood of every relevant image. The rank order of submissions was in the order that the user selected the relevant items, followed by the temporal neighbourhood images. In this way, the system maximised the potential for Recall, though at the expense of measures such as MAP.

NTCIR-14 Conference: Proceedings of the 14th NTCIR Conference on Evaluation of Information Access Technologies, June 10-13, 2019 Tokyo Japan



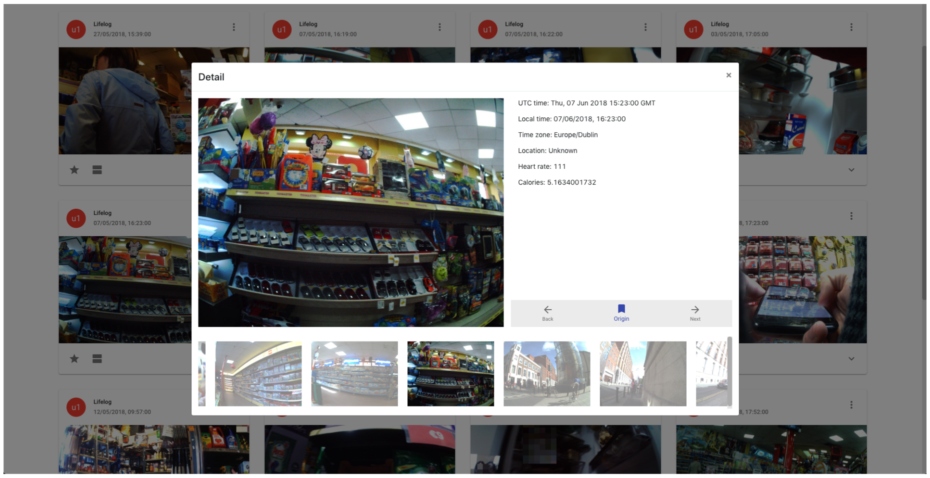


**Fig. 3.** Temporal Browsing

## 4 Interactive Experiment

In order to submit the official runs for the NTCIR14-Lifelog3 LSAT subtask, we organised an interactive user experiment in which novice users used the interactive retrieval system according to the following parameters and protocol.

### 4.1 Experimental Configuration

The evaluation was performed by four novice users whom each executed twelve topics. The topics were divided into two groups (1..12 and 13..24). Each user was given five minutes to complete each topic. The experimental protocol was as follows. The participant was introduced to the system and given a five minute review of functionality. Following this, the participant was allowed to test the system for a further ten minutes with two sample queries. Once the participant was comfortable with the system and how it operated, the user study began with the user reading a topic and the five minute timer was started once the user was comfortable that they understood the topic. All twelve topics were executed in forward order for users 1 and 2, and in reverse order for users 3 and 4. The whole experiment lasted about 90 minutes per user. In terms of practical experimental configuration, two users took part in the experiment in parallel (1 and 2, followed by 3 and 4).

### 4.2 Results

The User Experiment produced two runs; one combining the submissions of DCU-run1 (users 1, 2), and a second for DCU-run2(users 3, 4). DCU-run1 contained submitted results for 22 of the 24 topics, whereas DCU-run2 contained results for 23 of the topics. The number of retrieved relevant results for DCU-run1 was 556, whereas for DCU-run2, it was 1094. DCU-run2 found significantly more results that run 1, which highlights only a variability in how the teams were formed. Interestingly users 3 & 4 would have scored the system usability lower than users 1 & 2, although their interaction found over double the number of relevant items.

Considering that we were employing pagination of results at 20 per page, the P@10 metric for DCU-run1 was 0.1917 but for DCU-run2, it was 0.2292. Given the nature of the experiment, exploring results from a ranked list at higher cut-off points was not valuable. In terms of MAP, DCU-run1 was 0.0724, but for DCU-run2 it was 0.1274, once again significantly higher.

When comparing performance of this system with other participants in the LSAT sub-task (see table 1), it is apparent that the DCU-Run1 underperformed against other runs in terms of MAP and P@10, with only DCU-Run2 performing better than any competitor. It is our conjecture that this was due to the packing of the result submission with the temporal images, which would have reduced the MAP and P@10 scores. Considering the RelRet (Relevant items Recalled) measure, both runs were only bettered by the HCMUS system, which was the overall best performing interactive system. Once again, the submission packing would have increased these RelRet scores. Another factor that could be taken into consideration was the application of a five-minute time limit on each topic. Had this been longer, then the scores would likely have changed.

**Table 1.** Comparing DCU-Run1 & 2 with other Interactive Runs, from [2]

| Group ID | Run ID | Approach | MAP | P@10 | RelRet |
|---|---|---|---|---|---|
| NTU | NTU-Run2 | Interactive | 0.1108 | 0.3750 | 464 |
| NTU | NTU-Run3 | Interactive | 0.1657 | 0.6833 | 407 |
| HCMUS | HCMUS-Run1 | Interactive | 0.3993 | 0.7917 | 1444 |
| **DCU** | **DCURun1** | **Interactive** | **0.0724** | **0.1917** | **556** |
| **DCU** | **DCU-Run2** | **Interactive** | **0.1274** | **0.2292** | **1094** |

### 4.3 User Feedback

The inter-run comparisons just presented are not very useful when considering how well a system is liked by users. Clearly users 3 and 4 outperformed users 1 and 2. Using a questionnaire (The User Experience Questionnaire - QEU) [5], we sought to get an initial feedback from users about their experiences with the interactive retrieval engine. All four users filled in the simple 8 part questionnaire, which evaluated the system in terms of pragmatic (realistic-use-case) quality and hedonic (pleasantness) quality, with results shown in table 2. In terms of pragmatic quality, the interface was seen as being slightly more (+0.5 from a maximum of 3.0) supportive than obstructive, slightly more easy (+0.3) than complicated and slightly more clear (+0.3) than confusing. However users felt that it was slightly more inefficient (-0.3) than efficient. In terms of hedonic quality the interface was considered to be significantly more exciting (+1.3) than boring, significantly more interesting (+2.0) than non-interesting, significantly more inventive (+1.3) than conventional and slightly more leading-edge than usual/conventional.

**Table 2.** Pragmatic quality feedback of DCU's interactive retrieval engine

| | **Mean** | **Variance** | **Std. Dev.** | **Negative** | **Positive** | **Scale** |
|---|---|---|---|---|---|---|
| **1** | 0.5 | 6.3 | 2.5 | Obstructive | Supportive | Pragmatic Quality |
| **2** | 0.3 | 7.6 | 2.8 | Complicated | Easy | |
| **3** | -0.3 | 2.9 | 1.7 | Inefficient | Efficient | |
| **4** | 0.3 | 2.9 | 1.7 | Confusing | Clear | |
| **5** | 1.3 | 4.3 | 2.1 | Boring | Exciting | Hedonic Quality |
| **6** | 2.0 | 1.3 | 1.2 | Not Interesting | Interesting | |
| **7** | 1.3 | 2.3 | 1.5 | Conventional | Inventive | |
| **8** | 0.8 | 2.3 | 1.5 | Usual | Leading Edge | |

Exploring the qualitative findings on a per run basis, DCU-run1 users considered that the system was more supportive, easier to user, more efficient and clearer than DCU-run2 users. In terms of hedonic quality, they also found it more exciting, interesting, inventive and leading edge. However, considering the actual runs, these users were significantly less effective when using the system.

This feedback is reasonable because DCU-run2 users have prior experience of developing application system, which is why they expect the search engine to be more effective, clearer, and less complicated in interacting with our system. In contrast, DCU-run1 users understand how our search engine work after training without any further expectation of user interaction and think that our available functions are enough to retrieve the correct moments.

Through feedback and observation of the users using the retrieval system, we gathered findings that are being used to improve the current system for the LSC'19 (Lifelog Search Challenge) comparative benchmarking exercise. The new system called LifeSeeker [6] is an evolution of this system to incorporate the following updates:

- Taking measures to reduce the lexical gap (between user queries and the indexed concepts within the system) by expanding the indexed terms to include synonyms.
- Integrating content-similarity to allow the user to find similar looking content for any given image.
- Including a more conventional free-text search element and integrating the filter panel as part of the free-text query mechanism.
- Replacing pagination with lazy loading to show result fast and effectively.

## 5 Conclusions and Future Work

In this paper, we introduce a first-generation prototype of an interactive retrieval engine for lifelog data, that was run at the NTCIR14-Lifelog3 task. The system was a baseline retrieval system that operated over the provided metadata for the collection. The system was evaluated by four users and findings indicate that the system can be effectively used to locate relevant content. User studies showed that the users generally liked to system, but both observation and feedback provided a list of proposed enhancements to the system, which have been integrated into a new interactive retrieval system called LifeSeeker [6].

## Acknowledgements

This publication has emanated from research supported in party by research grants from Irish Research Council (IRC) under Grant Number GOIPG/2016/741 and Science Foundation Ireland under grant numbers SFI/12/RC/2289 and 13/RC/2106. We wish to also thank the four users who used our retrieval system to generate the offical runs.

## References

1. Duane, A., Gurrin, C., Huerst, W.: Virtual reality lifelog explorer: Lifelog search challenge at acm icmr 2018. In: Proceedings of the 2018 ACM Workshop on The Lifelog Search Challenge. pp. 20–23. LSC '18, ACM, New York, NY, USA (2018). https://doi.org/10.1145/3210539.3210544, http://doi.acm.org/10.1145/3210539.3210544
2. Gurrin, C., Joho, H., Hopfgartner, F., Dang-Nguyen, D.T., Zhou, L., Ninh, V.T., Le, T.K., Albatal, R., Healy, G.: Overview of NTCIR-14 lifelog task. In: Proceedings of the 14th NTCIR Conference on Evaluation of Information Access Technologies, National Center of Sciences, Tokyo, Japan, June 10-13, 2019 (2019)
3. Gurrin, C., Joho, H., Hopfgartner, F., Zhou, L., Albatal, R.: Overview of NTCIR-12 lifelog task. In: Proceedings of the 12th NTCIR Conference on Evaluation of Information Access Technologies, National Center of Sciences, Tokyo, Japan, June 7-10, 2016 (2016)
4. Gurrin, C., Schoeffmann, K., Joho, H., Münzer, B., Albatal, R., Hopfgartner, F., Zhou, L., Dang-Nguyen, D.T.: A Test Collection for Interactive Lifelog Retrieval. In: MultiMedia Modeling - 25th International Conference, {MMM} 2019, Thessaloniki, Greece, January 8-11, 2019, Proceedings, Part {I}. Lecture Notes in Computer Science, vol. 11295. Springer (2019)
5. Laugwitz, B., Held, T., Schrepp, M.: Construction and evaluation of a user experience questionnaire. In: Holzinger, A. (ed.) HCI and Usability for Education and Work. pp. 63–76. Springer Berlin Heidelberg, Berlin, Heidelberg (2008)
6. Le, T.K., Ninh, V.T., Dang-Nguyen, D.T., Tran, M.T., Zhou, L., Redondo, P., Smyth, S., Gurrin, C.: Lifeseeker - interactive lifelog search engine at lsc 2019. In: Proceedings of the 2019 ACM Workshop on The Lifelog Search Challenge. LSC '19, ACM, New York, NY, USA (2019)

7. Lin, T.Y., Maire, M., Belongie, S., Hays, J., Perona, P., Ramanan, D., Dollár, P., Zitnick, C.L.: Microsoft coco: Common objects in context pp. 740–755 (2014)
8. Lokoč, J., Souček, T., Kovalčik, G.: Using an interactive video retrieval tool for lifelog data. In: Proceedings of the 2018 ACM Workshop on The Lifelog Search Challenge. pp. 15–19. LSC '18, ACM, New York, NY, USA (2018). https://doi.org/10.1145/3210539.3210543, http://doi.acm.org/10.1145/3210539.3210543
9. Münzer, B., Leibetseder, A., Kletz, S., Primus, M.J., Schoeffmann, K.: lifexplore at the lifelog search challenge 2018. In: Proceedings of the 2018 ACM Workshop on The Lifelog Search Challenge. pp. 3–8. LSC '18, ACM, New York, NY, USA (2018). https://doi.org/10.1145/3210539.3210541, http://doi.acm.org/10.1145/3210539.3210541
10. de Oliveira Barra, G., Ayala, A.C., Bolaños, M., Dimiccoli, M., Giró i Nieto, X., Radeva, P.: Lemore: A lifelog engine for moments retrieval at the ntcir-lifelog LSAT task. In: Proceedings of the 12th NTCIR Conference on Evaluation of Information Access Technologies, National Center of Sciences, Tokyo, Japan, June 7-10, 2016 (2016)
11. Ren, S., He, K., Girshick, R., Sun, J.: Faster r-cnn: Towards real-time object detection with region proposal networks. In: Proceedings of the 28th International Conference on Neural Information Processing Systems - Volume 1. pp. 91–99. NIPS'15, MIT Press, Cambridge, MA, USA (2015), http://dl.acm.org/citation.cfm?id=2969239.2969250
12. Zhou, B., Lapedriza, A., Khosla, A., Oliva, A., Torralba, A.: Places: A 10 million image database for scene recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence (2017)