NTCIR-16 Lifelog-4 Guidelines

Cathal Gurrin, Frank Hopfgartner, Duc Tien Dang Nguyen
Hideo Joho, Thanh Binh Nguyen, Graham Healy, Liting Zhou

November 2021

v 1.0 - 12th Nov 2021

1 Introduction

NTCIR-Lifelog is a core task of the NTCIR-16 conference. This core task aims to advance the state-of-the-art research in lifelogging as an application of information retrieval. The methodology employed for the lifelog task at NTCIR-16 is based on the successfully deployed methodology from NTCIR-12-14.

For this year, we will focus on the most popular task from the previous lifelog tasks. 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.

2 Dataset Description

2.1 Overview

NTCIR16-Lifelog4 reuses an existing dataset, the LSC'21 dataset, which is a multimodal dataset that is four months in size, from one active lifelogger. The dataset consists of three files, each of which is password protected and details of how to access these datasets can be sourced from ntcir-lifelog@computing.dcu.ie with the competed agreement forms as descried below. The dataset comprises of:

- Core Image Dataset (38GB) of wearable camera images, fully redacted and anonymised in 1024 x 768 resolution, captured using OMG Autographer and Narrative Clip devices. These images were collected during periods in 2015, 2016 and 2018. All faces and readable text have been removed, as well as certain scenes and activities manually filtered out to respect local privacy requirements.

- Metadata for the collection (2.8MB), consisting of textual metadata representing time, physical activities, biometrics, locations, etc... Please note that there are no HR biometrics for the 2015 data.
- Visual Concepts (79.9MB) extracted from the non-redacted version of the visual dataset. Please see the ntcir website for details of the visual concept data.

Please note that the metadata, rather than the image filename, defines the time of capture. The metadata timezone is UTC time.

There will be 48 LSAT topics prepared for this task. 24 of the topics will be recall-focused, requiring as many potentially relevant items to be ranked as possible. The other 24 topics will be precision topics with only 1 or a small number of relevant items in the collection. Topics will be released in early December 2021.

Topics will follow the conventional TREC style, with title, description and narrative. An example topic is shown now:

```
TITLE: Preparing Meals
DESCRIPTION: Annotate all moments when the user is preparing meals.
NARRATIVE: Preparing meals involves the preparation of food items that occurs
before the user, or bystanders eat the meal. In order to be considered correct,
the food preparation process must be visible. The location where the food is
prepared does not matter.
```

3 Participation and Submission

In summary, each participating group should contact the organisers to get the collection used in the task. The required forms must be filled in as per the dataset instructions. When the dataset has been downloaded, then the participating team can index the content in whichever form is desired. There is no limitation on the type of data processing, enrichment or indexing process to be employed. Some participants may choose to index the provided metadata into a conventional inverted index, while others may choose to enhance the provided metadata using automated or semi-automated means, then index the data according to their preference.

Since there is only one subtask in NTCIR16-Lifelog4, then the participant must decide between two types of run to submit. Each run must be either automatic or interactive and must be appropriately labeled. The unit of retrieval (ranking) is the image ID (without JPG file extension).

- Automatic runs assume that there was no user involvement in the search process beyond specifying the initial query, which can only happen once for each topic. The search system generates a ranked list of up to 100 images for each topic. There is no time limit on how long it can take for an automatic run. We assume that any human involvement in generating a query from the topic is a once-off process that is not iterative and dependent on the results of a previous execution of a query for that topic (i.e. no human-influenced relevance feedback mechanism can be applied to an automatic run). The submission file format includes SCORE to capture the score of every image as returned by the ranking algorithm. In the automatic

run case the SECONDS-ELAPSED column should always have a value equal to 0, since it is only relevant for Interactive runs.

- Interactive runs assume that there is a user involved in the search process that generates a query and selects which moments are considered correct for each topic. This may be a single phase, or may contain multiple phases of relevance feedback or query reformulation. In interactive runs, the maximum time allowed for any topic should be 300 seconds. The submission file format includes SECONDS-ELAPSED to capture the time taken to find every moment. In the interactive runs, mo more than 100 images may be submitted for any query also. In the interactive run, the SCORE value should be equal to 1. For interactive runs, the seconds elapsed should be equal to the number of seconds (from zero) that it took the user to find a particular item in the submission. For example, if a user in an interactive run found one potentially relevant item at 5 seconds, another at 15 seconds and a third at 255 seconds, then there would be three lines in the CSV file for that run, each of which has a different value for the SECONDS-ELAPSED column. It is important to accurately record this value since it will be used to calculate run performance at different time cutoffs (e.g. 10 seconds, 60 seconds, etc...).

A submitted run for the LSAT task is in the form of a single CSV file per run. Please note that each group can submit up to 10 runs, each as an individual file. The submission files should be sent (one email per group) to (ntcir-lifelog@computing.dcu.ie) by the due date with the title 'NTCIR-Lifelog LSAT Submission'. The submission file should be named as follows: GroupID-RunID-[Interactive or Automatic].txt, where GroupID is the registration ID of your group at NTCIR, RunID is the number of the run (e.g. DCULSAT01 or DCULSAT02, etc..), and the label Automatic or Interactive.

For every topic, every image considered relevant should have one line in the CSV file. For some topics there will be only one relevant item (one line in the file), for others there will be many relevant items (many lines in the file), up to 100. It is also possible that no relevant items are found for a topic, so then there should be no entry in the file for the topic.

The format of the CSV file for an automatic run would be as follows:

```
GROUP-ID, RUN-ID, TOPIC-ID, IMAGE-ID, SECONDS-ELAPSED, SCORE
...
DCU, DCULSAT01, LSAT003, u1_2016-08-15_112559, 0, 1.0
DCU, DCULSAT01, LSAT003, u1_2016-08-15_120354, 0, 1.0
...
```

The format of the CSV file for an interactive run would be as follows:

```
GROUP-ID, RUN-ID, TOPIC-ID, IMAGE-ID, SECONDS-ELAPSED, SCORE
...
DCU, DCULSAT01, LSAT003, u1_2016-08-15_112559, 33, 1.0
DCU, DCULSAT01, LSAT003, u1_2016-08-15_120354, 54, 1.0
DCU, DCULSAT01, LSAT003, u1_2016-08-15_120412, 243, 1.0
...
```

The trec-eval programme [1] will be employed to generate result scores for each run. Relevance judgements will be generated using a pooled approach whereby human judges will be employed to

manually evaluate each submitted image for each topic, up to a maximum of 100 images per topic, per run, per participant. The relevance judgements will be binary and will be informed by the immediate context of each image if important.

Following submission, each participating team must prepare a paper describing their experimental approach and scores. The organisers will prepare their own Overview Paper [2] which should be referenced by all participants.

References

[1] NIST. Trec-eval. https://trec.nist.gov/trec$_eval$/, 2022($accessedNovember21, 2022$).

[2] Cathal Gurrin, Frank Hopfgartner, Duc-Tien Dang-Nguyen, Thanh-Binh Nguyen, Graham Healy, Rami Albatal, and Liting Zhou. 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, 2022.