

NTCIR12-Lifelog, A Test Collection to Support Collaborative Benchmarking of Lifelog Retrieval Systems

Liting Zhou¹, Hideo Joho², Frank Hopfgartner³, Rami Albatal¹, Cathal Gurrin¹

¹Insight Centre for Data Analytics at Dublin City University, Ireland.

²Tsukuba University, Japan

³University of Glasgow, United Kingdom

Abstract

Test collections have a long history of supporting repeatable and comparable evaluation in Information Retrieval. In this paper we describe NTCIR12-Lifelog, a first lifelog test collection to support comparative benchmarking of lifelog retrieval systems. We provide an overview summary of the test collection and motivate use-cases of its application. This test collection is being employed at NTCIR-12 to evaluate approaches to Lifelog data search and organisation.

Keywords: lifelogging; test collection; privacy-by-design; personal data

Citation: Editor will add citation

Copyright: Copyright is held by the authors.

Acknowledgements: This research was supported by Science Foundation Ireland under grant number SFI/12/RC/2289.

Contact: becky.zhou@dcu.ie.

1 Introduction

For many years, test collections have been used to support repeatable and comparative evaluation in Information Retrieval and related fields. The earliest example is the pioneering Cranfield collection of 1,398 abstracts of aerodynamics journal articles, a set of 225 queries, and exhaustive relevance judgments of all (query, document) pairs, which was gathered in the late 1950s to support early experimentation into Information Retrieval (Van Rijsbergen & Sparck Jones, 1972). Since then the field has embraced the concept of test collections for reasons such as:

- Lowering the cost of experimentation by removing the need for researchers to gather documents and construct relevance judgements, and;
- Helping to ensure that experiments are both repeatable and comparable across sites and time.

Consequently, the generation of test collections has been a primary focus of researchers in all aspects of Information Retrieval and related disciplines. The process has been formalised under the umbrella of a number of organisations such as TREC in the US, CLEF in Europe and NTCIR in Asia.

One aspect of Information Retrieval that has been gathering increasing attention in recent years is the concept of lifelogging. Lifelogging is defined as “a form of pervasive computing, consisting of a unified digital record of the totality of an individual’s experiences, captured multi-modally through digital sensors and stored permanently as a personal multimedia archive” (Gurrin, Smeaton & Doherty, 2014). Lifelogging typically generates multimedia archives of life-experience data in an enormous (potentially multi-decade) lifelog. Such a lifelog needs to be organised and searchable to be valuable to the lifelogger. Hence there have been calls for a test collection of lifelog data. The contribution of this paper is an overview description of the NTCIR12-Lifelog test collection.

2 A Description of the NTCIR-12 Lifelogging Test Collection

The NTCIR12-Lifelog test collection was created in 2015 as a multimodal dataset of three months of real-world lifelog data. Accompanying this dataset was a set of topics and relevance judgements supporting ad-hoc style retrieval.

The NTCIR Lifelog dataset consists of a subset of a real-world lifelog from three lifeloggers for a period of about one month each. The lifeloggers were asked to gather data during as much of their waking hours as possible, and typically consisted of all-day data. The data consists of a large collection of 88,124

wearable camera images (from the OMG Autographer camera) and anonymised via the process previously described. An example of such images shown in Figure 1. The dataset also contained It also contains semantic locations (e.g. Starbucks cafe, McDonalds restaurant, home, work), physical activities (e.g. walking, transport, cycling) and an XML description of this data.

Component	Value
Number of Lifeloggers	3
Size of the Collection (Days)	87 days
Size of the Collection (Images)	88,124 images
Size of the Collection (GB)	18.18 GB
Size of the Collection (Number of Long-stay Semantic Locations)	130 locations
Size of the Collection (Visual Concept Metadata)	840 MB
Number of LSAT (Ad-hoc) Topics	48
Number of Insight Topics	10

Table 1. A Basic Analysis of the NTCIR12-Lifelog Test Collection

Given the fact that lifelog data is typically visual in nature, the output of the CAFFE CNN-based visual concept detector (Jia et al, 2014) was included in the test collection as additional metadata. This classifier provided labels and probabilities of occurrence for 1,000 objects in every image. A summary of the test collection is shown in Table 1 and sample images in Figure 1. The data was roughly equally distributed per lifelogger.



Figure 1. Sample Wearable Camera Images from the NTCIR12-Lifelog Test Collection

Aside from the data, the test collection includes a set of topics (queries) that are representative of the real-world information needs of lifeloggers (Sellen & Whittaker, 2010). There are 48 ad-hoc search topics representing the challenge of *Retrieval* from memories. These topics were generated by the three lifeloggers and represent real information needs that they would have had for their lifelog data. The relevance judgements for NTCIR12-Lifelog were manual (non-pooled) relevance judgements, and were generated by the lifeloggers and data organisers for all 48 LSAT topics. These are used to compare the participant submissions and form the third component of the NTCIR Lifelog test collection that supports comparative and repeatable experimentation into organisation and search of lifelog archives.

3 References

- Gurrin, C., Smeaton, A.F & Doherty, A.R. (2014). *Lifelogging: Personal Big Data*. (1st ed.). Hanover, MA, USA: Now Publishers Inc.
- Jia et al.. (2014). Caffe: Convolutional Architecture for Fast Feature Embedding. *MM '14 Proceedings of the 22nd ACM international conference on Multimedia*, New York, USA., 675-678.
- Sellen, A & Whittaker, S. (2010). Beyond total capture: a constructive critique of lifelogging. *Communications of the ACM*, 53(5), 70-77.
- Van Rijsbergen, C. .J. & Sparck Jones, K. (1972). A Test for the Separation of Relevant and Non-Relevant Documents in Experimental Retrieval Collections. *Journal of Documentation*, 29(3).