Big Data For Lifelog Moments Retrieval Improvement

Fatma Ben Abdallah[1,2], Ghada Feki[1], Anis Ben Ammar[1], and Chokri Ben Amar[1]

[1] REGIM-Lab.: REsearch Groups in Intelligent Machines, University of Sfax, National Engineering School of Sfax (ENIS), BP 1173, Sfax, 3038, Tunisia
{ben.abdallah.fatma,ghada.feki,anis.ben.ammar,chokri.benamar}@ieee.org
http://www.regim.org/

[2] Higher Institute of Technological Studies, ISET Kairouan, 3199 Raccada, Tunisia
http://www.isetkr.rnu.tn/

Abstract. The ImageCLEF lifelog Moment Retrieval Task promotes research for lifelogging retrieval by providing a benchmark with an evaluation process that allows a comparative analysis of lifelog methods, approaches and tools. In this paper, we describe our participation at the ImageCLEF lifelog LMRT 2019. Findings from our initial experiments in the LMRT sub-task in ImageCLEFlifelog2018 have motivated us to improve our deep learning-based processing for lifelog image retrieval approach using NoSQL database. The new version employs a distributed database and framework for storing and processing large volumes of data. We try to reduce user involvement during the fine-tuning phase by using the ground truth for the development dataset. We implement our architecture using Matlab, Cassandra, and Spark. The best results were given by the first run with precision@10=0.28. This run is based on fine-tuning Googlenet with the weights freeze of the 110 first layers.

Keywords: Deep-learning · Transfer-learning · Big Data · NoSQL · Lifelog · Moments Retrieval.

1 Introduction

Lifelogging is a concept that has emerged in recent years to translate people's interest in the daily logging of their lives. The lifelogging is intended for private use, unlike social networks which are also kinds of lifelogs. Many devices and applications are available to monitor our training, diet, health, sleep, etc. The information collected by these devices and applications is characterized by the heterogeneity and the multimodality which make the research process in this mass of data a complex and non-trivial task. Considering this context, several workshops, panels, and evaluation campaign offers research tasks to deal with the

problem of retrieving, summarizing and visualizing lifelogging data. Given the huge amount of digital data, it has become necessary to develop new methods to manage and analyze them. Big Data is about finding, capturing, storing, sharing and presenting this data. To satisfy the need of heavy data processing, NoSQL is a database query language for Big Data. Several NoSQL database exist with document store (CouchDB, MongoDB, TerraStore, eXist, Virtuoso), key-value store (DynamoDB, Voldemort, Azure Table Storage, MongoDB), graph store (AllegroGraph, InfiniteGraph) and, tabular store (Cassandra, Hadoop / Hbase, Hypertable). Since that storing into tables allows greater ease of developpement with a SQL-like language with CQL, our choice focused on Apache Cassandra.
Our research to date has focused on proposing a deep learning-based processing approach for lifelog image retrieval [2–4]. Compared to our initial approach [3], our participation in the ImageCLEF Lifelog Moment Retrieval Task 2019 (LMRT) [8] which is part of the Conference and Labs of the Evaluation Forum (CLEF 2019) [9] has two main improvements. First, we use the ground truth of the development dataset to automatically dispatch images into categories for the fine-tuning. Second, we use Apache Cassandra a NoSQL-based database management system (DBMS) designed to handle massive amounts of data. Cassandra Query Language (CQL) only implements a subset of SQL, so we use Spark with Cassandra to operate data analytics that CQL doesn't provide. From an initial query, our approach can automatically extract from it relevant concepts based on Long-Term-Short-Memory(LSTM). After that, the retrieval phase consists in searching the extracted query concepts in the file containing the image concepts. The runs submitted in the LMRT 2019 vary in the generation of image concepts. For the first run, we fine-tune Googlenet with the weight freeze of the first 110 layers. For the second run, we fine-tune Googlenet without freezing. For the third run, we fine-tune Alexnet. For the fourth and fifth run, we use respectively Googlenet and Alexnet to classify all the images of the test dataset. For the sixth run, we used only the textual features given by the organizers. The best results were given by the first run with F1-measure=0.188, ranked sixth in the challenge.
The remainder of this paper is divided into five sections. In section 2, we present existing retrieval architectures using fine-tuning. In section 3, we detail our approach. Section 4 presents the experimental results of our implementation. Section 5 provides some concluding remarks and suggests future works

2 Related Work

Training models to have human-like capabilities requires a lot of resources in terms of data and time. To optimize this learning, we must pool knowledge from one model to another by practicing transfer learning and especially fine-tuning. Mainly, in the case of image processing, we re-use the layers of a model that has already learned that we fixed, the last layers as for it will change according to the learning data and will refine according to input data.
Babenko et al. [1] have demonstrated a significant improvement in the search

performance of the neural network when it is trained on a dataset which is similar to the one encountered during the test phase. Bases of this observation, we focused our study on existing retrieval architectures using fine-tuning.
Authors in [13] proposed a CNN framework for clothes image retrieval in recommendation system. The first framework's module use Alexnet pre-trained on Imagenet for learning rich mid-level visual representations. The second one, fine-tune the Alexnet network on clothing dataset using backpropagation. Finally, the images are retrieved via hierarchical deep search. Authors in [14] proposed to fine-tune CNN for image retrieval from a large collection of images using 3D reconstruction and siamese architecture. In [5], the authors investigate the use of CNN-based features for food retrieval. They use the last fully connected layer of Resnet-50 as a feature extractor. The most similar approach to our lifelog image retrieval context is proposed in [17]. The authors developed a general framework to translate lifelog images into features. They choose to fine-tune VGG-16 pre-trained on ImageNet1K[3] by replacing the last layer which contains 1000 neurons with 634, followed by sigmoid activation instead of softmax function. Other frameworks/systems for lifelog image retrieval were proposed in [16, 18–20] and rely only on CNN and DNN pre-trained on Imagenet to extract feature. Fine-tuning method outperforms those use only pre-trained CNN on Imagenet1K, Places365[4] or MSCOCO[5]. These affirmations are confirmed by the experimental results conducted in section 4.

3 Proposed Approach: Big Data For Lifelog Moments Retrieval

3.1 Overview

We based our proposed approach on our previous participation in LMRT 2018 [7]. However, the main difference between the two approaches is that the new version employs a distributed database and framework for storing and processing large volumes of data. Furthermore, we automate the images dispatching for the fine-tuning phase by using the ground truth for the development dataset. The ImageCLEF LMRT dataset, in addition to containing images, includes a set of metadata consisting of biometrics, historic glucose index, semantic locations visited, physical activities, attributes predicted by using the Place CNN trained on SUNattribute dataset and also trained on Place 365 dataset, the class name, the bounding box and the score of the 25 objects with the highest score in each image predicted by using Faster R-CNN [15] trained on the COCO dataset. To exploit these huge lifelog metadata description, we use CQL Query with Spark connector. From an initial query, our approach can automatically extract from its relevant concepts based on Long-Term-Short-Memory(LSTM). After that, the retrieval phase consists in searching the extracted query concepts in the file

[3] http://www.image-net.org/
[4] http://places2.csail.mit.edu/download.html
[5] http://cocodataset.org/

containing the image concepts.
Fig. 1 presents the overview of our proposed approach for lifelog moments retrieval. Our approach has five main phases:

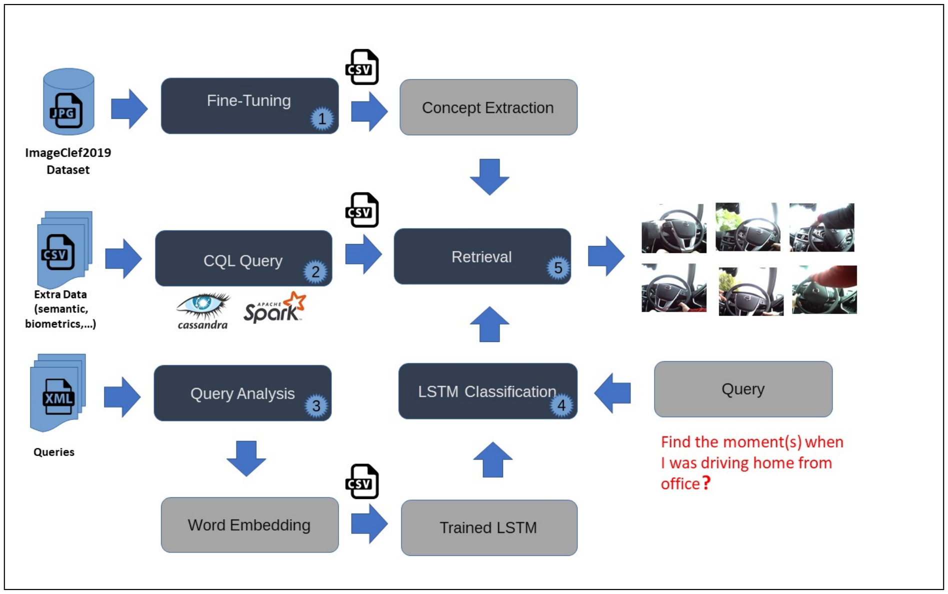

Fig. 1. Proposed architecture

– **Fine-Tuning**: We organize 9676 images into 51 classes using the ground truth (GT) of the development set. The classes were formed from the clusters file provided with the GT. For the clusters restaurant, watch videos and cooking which contain several sub-clusters (respectively 11, 31 and 10), we had to check visually the images to more appropriately rename the classes. Nevertheless, several topics do not appear in the GT. For the topic in a toyshop and seeking food in the fridge, we choose manually the images. For the topic coffee time, we used the images from the imageCLEF LMRT 2018. For the CNN parameters, we used the same setting as last participation last year[3]. In fact, we replace the last three layers of the network: a fully connected layer, a softmax layer, and a classification output layer. Besides, during the training process with 70% of the images for training and 30% for validation, we use data augmentation to prevent the network from overfitting. After the training, we classify all the images from the dataset and generate CSV file containing for each image the concept with the highest score.

– **CQL Query**: For each lifelogger u1 and u2, two tables were provided: the minute-based table and the categories and concepts table. We import for each lifelogger a table in the DBMS Cassandra, then we use Scala on Spark to write CQL query. By analyzing topics in the test set, we find that each one can be divided into 5 axes to facilitate the retrieval process : user, concept, activity, location, and irrelevance. These 5 axes can be translate to a CQL query : *Select column from user table where condition*. The column contains the activity, the concepts and the location. In the condition we can use «NOT IN» to express irrelevance, «IN», «CONTAINS» or «LIKE» to express matching with concepts. The possibilities are numerous.

- **Query Analysis**: To extract relevant concepts from the given query, we build labeled textual descriptions of queries moments. For that purpose, we used the development and the test set topics of the NTCIR-12 [11], NTCIR-13, imageCLEF LRT 2017 [6] and imageCLEF LMRT 2018 that we have combined. We obtained a csv file which contains for each topic title, the topic description and the relevant concepts associated with the topic. We convert the concepts to numeric vectors by training a word embedding. After that, we create and train an LSTM network based on the sequences of word vectors. The concepts used in query analysis phase are the same as the one defined in fine-tuning phase.

- **LSTM Classification**: We choose to use LSTM in our architecture to predict concepts for any given user query and not only those given by the organizers. We aim to be general and global. For example, we take the case of a lifelogger searching for the moments who shows him driving. He writes either the title or the description and the trained LSTM will return the concepts «steering wheel, windshield».

- **Retrieval**: The retrieval phase consists of matching the extracted concepts from the LSTM with the concept from the first phase of fine-tuning. For example, with the concepts «steering wheel, windshield» gived from the previous LSTM classification phase, we perform a simple matching of these concepts in the CSV file generated during the first phase. After that, we sort decreasingly the result to obtain the highest score values.

4 Results obtained

We submitted 6 runs on the LMRT subtask 2019 summarized in Table 1.

The runs submitted in the LMRT 2019 vary in the generation of image concepts. For the first run, we fine-tune Googlenet with the weight freeze of the first 110 layers. For the second run, we fine-tune Googlenet without freezing. For the third run, we fine-tune Alexnet. For the fourth and fifth run, we use respectively Googlenet and Alexnet to classify all the images of the test dataset. For the sixth run, we used only the textual features given by the organizers. Fig. 2 presents the detailed results of all teams that participated to the ImageCLEF LMRT tasks. Official ranking metrics is the F1-measure@10, which gives equal importance

Table 1. Submitted Runs

Run	Name	Parsing	Type of information
RUN_1	Fine-tuning Googlenet with freeze	Automatic	Visual
RUN_2	Fine-tuning Googlenet without freeze	Automatic	Visual
RUN_3	Fine-tuning with Alexnet	Automatic	Visual
RUN_4	Classifying with Googlenet	Automatic	Visual
RUN_5	Classifying with Alexnet	Automatic	Visual
RUN_6	Baseline : CQL on organizers concepts	Automatic	Textual

to diversity (via CR@10) and relevance (via P@10). The best team HCMUS obtained F1-measure@10=0.61 with an interactive approach [12]. For our team REGIMLAB, which proposed only automatic approach, the best results were given by the first run with F1-measure=0.188, ranked sixth in the challenge. The results of the runs submitted to the LMRT 2019 subtask are detailed in tables 2, 3 and 4.

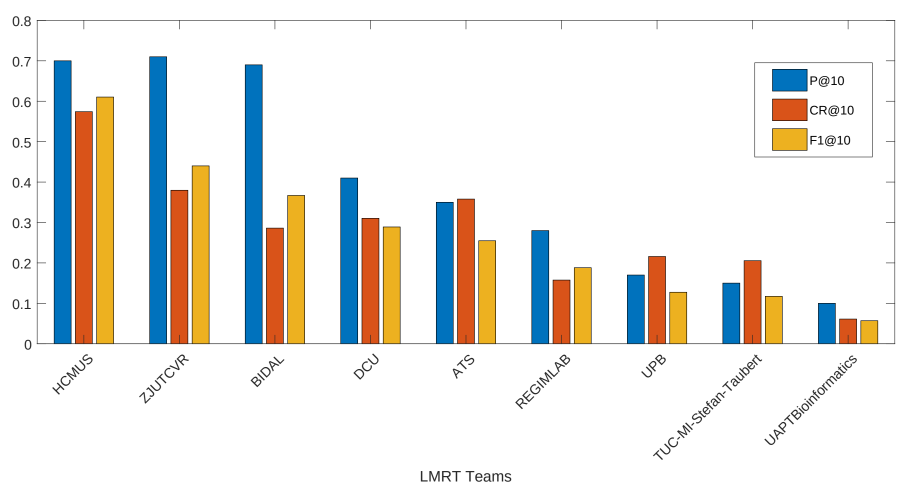

Fig. 2. ImageCLEF LMRT Official results

5 Analysis of the results

By analyzing F1-measure@10 results for each query Fig.3, we note that a neural network that is trained on a dataset which is similar to the one encountered during the test phase demonstrated a significant improvement in the search performance. Furthermore, fine-tuning with Googlenet with freezing the weights of

Table 2. Precision at X (P@X)

Cut-off	P@5	P@10	P@20	P@30	P@40	P@50
RUN_1	0.280	**0.280**	0.240	0.237	0.233	0.222
RUN_2	0.260	0.250	0.235	0.220	0.218	0.214
RUN_3	0.260	0.250	0.240	0.227	0.223	0.220
RUN_4	0.100	0.090	0.100	0.103	0.093	0.088
RUN_5	0.080	0.070	0.070	0.067	0.063	0.064
RUN_6	0.060	0.070	0.055	0.040	0.043	0.040

Table 3. Cluster Recall at X (CR@X)

Cut-off	CR@5	CR@10	CR@20	CR@30	CR@40	CR@50
RUN_1	0.136	0.158	0.179	0.189	0.225	0.230
RUN_2	0.125	0.142	0.153	0.186	0.222	0.222
RUN_3	0.092	0.103	0.197	0.219	0.250	**0.256**
RUN_4	0.026	0.048	0.142	0.159	0.159	0.170
RUN_5	0.072	0.088	0.104	0.120	0.136	0.136
RUN_6	0.055	0.076	0.087	0.087	0.103	0.125

Table 4. F1-measure at X (F1@X)

Cut-off	F1@5	F1@10	F1@20	F1@30	F1@40	F1@50
RUN_1	0.168	0.188	0.180	0.180	**0.207**	0.202
RUN_2	0.153	0.167	0.161	0.170	0.197	0.194
RUN_3	0.133	0.142	0.185	0.183	0.191	0.190
RUN_4	0.041	0.062	0.096	0.111	0.100	0.101
RUN_5	0.060	0.062	0.069	0.074	0.076	0.073
RUN_6	0.037	0.061	0.054	0.045	0.052	0.055

the 110 first layers gave better performance than with Alexnet.
We also see that for the second topic «*Find any moment when u1 was driving home from the office*», the run 6 which is based on CQL overpass the best run. This is due to the consideration of the latitude and longitude of volunteer's position described in the lifelog metadata description.
Considering the precision measure in Fig.4 which assess the proportion of relevant documents found among all documents found by the system, we can notice a considerable difference from one query to another. The accuracy depends essentially on the examples that were provided during the transfer learning. The three first runs that are based on fine-tuning achieved a precision@10=1 for the query : «*Find the moment when either u1 or u2 was watching football on the TV*». Besides, despite the fact that the fine-tuned images for the class coffee come from another lifelogger, we achieved a precision@10=0.7 for the query :«Find the moment when u1 was having coffee in a cafe ».

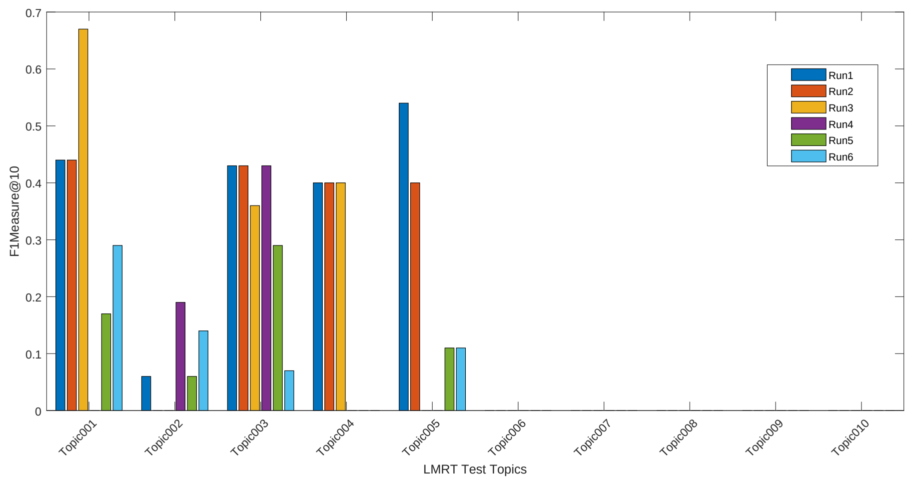

Fig. 3. F1-measure@10 scores on the test set

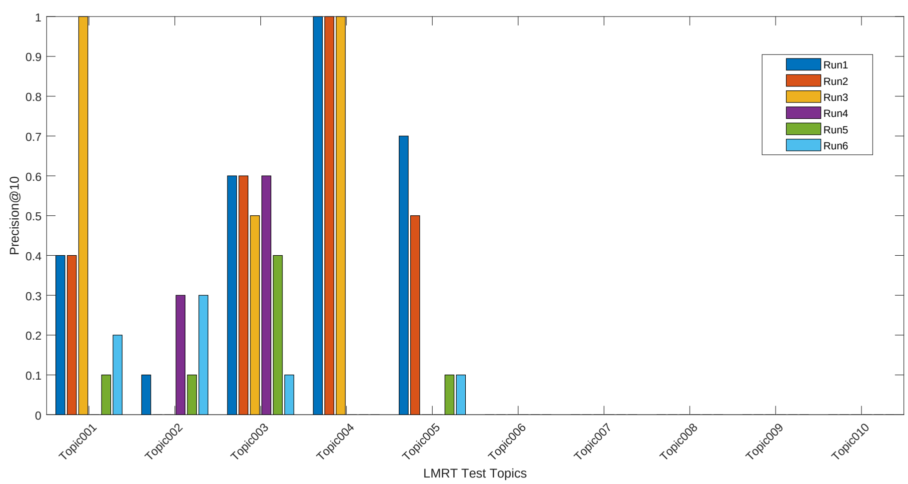

Fig. 4. Precision@10 scores on the test set

6 Conclusion and perspectives for future work

This paper presents our approach for lifelog moment retrieval at the ImageCLEF Lifelog Moment Retrieval Task 2019. The new version, compared to the previous participation at the LMRT 2018, employs a distributed database and framework for storing and processing large volumes of data. The best performance was reached by an interactive approach which incite us to include the user in the

process.
The time limit and technical problems did not allow us to submit all the planned runs. So as future work, we will combine visual and textual features to improve the results. We also plan to perform neural network training on more powerful computers with more GPUs to reduce learning time. Besides, we should filter the dataset before the fine-tuning by removing uninformative and blurry images.

7 Acknowledgments

The research leading to these results has received funding from the Ministry of Higher Education and Scientific Research of Tunisia under the grant agreement number LR11ES48.

References

1. Babenko, A.; Slesarev, A.; Chigorin, A. & Lempitsky, V. Fleet; D., Pajdla; T., Schiele, B. & Tuytelaars, T. (Eds.), Neural Codes for Image Retrieval Computer Vision, ECCV , Springer International Publishing, pp. 584-599 (2014)
2. Abdallah, F. B.; Feki, G.; Ammar, A. B.; & Amar, C. B., A new model driven architecture for deep learning-based multimodal lifelog retrieval. In: poster Proceedings: International Conference WSCG, CSRN 2803, pp 8-17, Czech Republic(2018)
3. Abdallah, F.B.; Feki, G.; Ezzarka, M.; Ammar, A.B. & Amar, C.B., Regim Lab Team at ImageCLEF Lifelog Moment Retrieval Task 2018. Working Notes of CLEF 2018, France(2018)
4. Abdallah, F. B.; Feki, G.; Ammar, A. B. & Amar, C. B., Multilevel Deep Learning-Based Processing for Lifelog Image Retrieval Enhancement. In: Proceedings of the IEEE International Conference SMC, pp 1348-1354, Japan(2018)
5. Ciocca, G.; Napoletano, P. & Schettini, R. Battiato, S.; Farinella, G. M.; Leo, M. & Gallo, G. (Eds.), Learning CNN-based Features for Retrieval of Food Images, New Trends in Image Analysis and Processing, ICIAP, Springer International Publishing, pp. 426-434 (2017)
6. Dang-Nguyen, D.-T.; Piras, L.; Riegler, M.; Boato, G.; Zhou, L. & Gurrin, C., Overview of ImageCLEFlifelog 2017: Lifelog Retrieval and Summarization CLEF2017 Working Notes, CEUR-WS, 2017, 1866
7. Dang-Nguyen, D.-T.; Piras, L.; Riegler, M.; Zhou, L.; Lux, M. & Gurrin, C., Overview of ImageCLEFlifelog 2018: Daily Living Understanding and Lifelog Moment Retrieval CLEF2018 Working Notes, CEUR-WS.org /ceur-ws.org>, 2018
8. Dang-Nguyen, D.-T.; Piras, L.; Riegler, M.; Tran, M.-T.; Zhou, L.; Lux, M.; Le, T.-K.; Ninh, V.-T. & Gurrin, C., Overview of ImageCLEFlifelog 2019: Solve my life puzzle and Lifelog Moment Retrieval CLEF2019 Working Notes, CEUR-WS.org , 2019
9. Ionescu, B.; Müller, H.; Péteri, R.; Cid, Y. D.; Liauchuk, V.; Kovalev, V.; Klimuk, D.; Tarasau, A.; Abacha, A. B.; Hasan, S. A.; Datla, V.; Liu, J.;

Demner-Fushman, D.; Dang-Nguyen, D.-T.; Piras, L.; Riegler, M.; Tran, M.-T.; Lux, M.; Gurrin, C.; Pelka, O.; Friedrich, C. M.; de Herrera, A. G. S.; Garcia, N.; Kavallieratou, E.; del Blanco, C. R.; Rodríguez, C. C.; Vasillopoulos, N.; Karampidis, K.; Chamberlain, J.; Clark, A. & Campello, A., ImageCLEF 2019: Multimedia Retrieval in Medicine, Lifelogging, Security and Nature Experimental IR Meets Multilinguality, Multimodality, and Interaction, LNCS Lecture Notes in Computer Science, Springer (2019)

10. Gurrin, C.; Joho, H.; Hopfgartner, F.; Zhou, L.; Gupta, R.; Albatal, R.; Nguyen, D.T.D., Overview of NTCIR-13 Lifelog-2 Task. Proceedings of the Thirteenth NTCIR conference (NTCIR-13), Tokyo, Japan (2017)
11. Gurrin, C.; Joho, H.; Hopfgartner, F.; Zhou, L.; & Albatal, R. (2016)., Overview of NTCIR-12 Lifelog Task. Proceedings of the 12th NTCIR Conference on Evaluation of Information Access Technologies, Japan (2016)
12. Le, N.; Nguyen, D.; Hoang, T.; Nguyen, T.; Truong, T.; Duy, T. D.; Luong, Q.; Vo-Ho, V.; Nguyen, V. & Tran, M., HCMUS at the NTCIR-14 Lifelog-3 Task. Proceedings of the 14th NTCIR Conference on Evaluation of Information Access Technologies, Japan (2019)
13. Lin, K.; Yang, H.-F.; Liu, K.-H.; Hsiao, J.-H. & Chen, C.-S. Rapid Clothing Retrieval via Deep Learning of Binary Codes and Hierarchical Search Proceedings of the 5th ACM on International Conference on Multimedia Retrieval, pp. 499-502, ACM (2015)
14. Radenovic, F.; Tolias, G. & Chum, O. Leibe, B.; Matas, J.; Sebe, N. & Welling, M. (Eds.) CNN Image Retrieval Learns from BoW : Unsupervised Fine-Tuning with Hard Examples Computer Vision, ECCV, Springer International Publishing, pp. 3-20 (2016)
15. Ren, S.; He, K.; Girshick, R. & Sun, J. Faster r-cnn: Towards real-time object detection with region proposal networks Advances in neural information processing systems, 2015, 91-99
16. Safadi, B.; Mulhem, P.; Qunot, G.; Chevallet, J.P. : LIG-MRIM at NTCIR-12 Lifelog Semantic Access Task, In The 12th NTCIR Conference, Tokyo, Japan (2016)
17. Lin, J.; del Molino, A. G.; Xu, Q.; Fang, F.; Subbaraju, V.; Lim, J. H.; & Chandrasekhar, V., VCI2R at the NTCIR-13 Lifelog-2 Lifelog Semantic Access Task. In : Thirteenth NTCIR conference (NTCIR-13), Tokyo, Japan (2017)
18. Xia L.; Ma Y.; Fan W. : VTIR at the NTCIR-12 2016 Lifelog Semantic Access Task. In The 12th NTCIR Conference, Evaluation of Information Access Technologies, Tokyo, Japan (2016)
19. Yamamoto, S.; Nishimura, T.; Akagi, Y.; Takimoto, Y.;Inoue, T. & Toda, H. Pbg at the ntcir-13 lifelog-2 lat, lsat, and lest tasks, Proceedings of NTCIR-13, Tokyo, Japan, (2017)
20. Zhou, L.; Piras, L.; Riegler, M.; Boato, G.; Nguyen, D.T.D.; Gurrin C., Organizer Team at ImageCLEFlifelog 2017: Baseline Approaches for Lifelog Retrieval and Summarization, In : CLEF2017 Working Notes (CEUR Workshop Proceedings), Dublin, Ireland (2017).