VieLens, An Interactive Search engine for LSC2019

Son P. Nguyen
Dien H. Le
Uyen H. Pham
VNU HCM University of Economics and Law
Vietnam

Martin Crane
Graham Healy
Cathal Gurrin
Dublin City University
Ireland

ABSTRACT

With the appearance of many wearable devices like smartwatches, recording glasses (such as Google glass), smart phones, digital personal profiles have become more readily available nowadays. However, searching and navigating these multi-source, multi-modal, and often unstructured data to extract useful information is still a relatively challenging task. Therefore, the LSC2019 competition has been organized so that researchers can demonstrate novel search engines, as well as exchange ideas and collaborate on these types of problems. We present in this paper our approach for supporting interactive searches of lifelog data by employing a new retrieval system called *VieLens*, which is an interactive retrieval system enhanced by natural language processing techniques to extend and improve search results mainly in the context of a user's activities in their daily life.

CCS CONCEPTS

• **Information systems → Search interfaces**; *Multimedia databases*; • **Human-centered computing** → Interactive systems and tools.

KEYWORDS

datasets, lifelog, interactive retrieval, natural language processing, machine learning

ACM Reference Format:
Son P. Nguyen, Dien H. Le, Uyen H. Pham, Martin Crane, Graham Healy, and Cathal Gurrin. 2019. VieLens, An Interactive Search engine for LSC2019. In *Lifelog Search Challenge'19 (LSC'19), June 10–13, 2019, Ottawa, ON, Canada.* ACM, New York, NY, USA, 3 pages. https://doi.org/10.1145/3326460.3329161

1 INTRODUCTION

Lifelogging was popularised by Gordon Bell with the "MyLifeBits" project for the purpose of recording every aspect of his life digitally [6]. Nowadays, with devices like smartwatches, Google glass etc., it is a lot more convenient to lifelog all aspects of daily life, including activities like eating meals, going shopping, hanging out, as well as gathering detailed health and biometric data such as heart rate, blood pressure and so on. Thus the focus of attention moves has moved largely from collection of data to deriving information from it.

There have been a variety of applications of lifelogging data, especially in memory support [5], monitoring diet and health [9], [1], and in large-scale epidemiology studies [12]. For all these applications, the user interface is perhaps the most important component of the information management system, because the users of the system are unlikely to be experts in either HCI or interactive retrieval system development. In this paper, we propose a method of searching the lifelogging data with an emphasis on activities of daily living (ADLs), with a special focus on the activities of a consumer in daily life, due to the requirements of the project that funds this research.

In searching for activities in lifelogging data, a number of different techniques have been used, from simple keyword filters to sophisticated image analysis techniques. In this paper, we present results from a search engine which we have created incorporating a state-of-the-art algorithm in natural language processing to extend the list of keywords or phrases provided by users. Our observation is that the automatic labeling of images is not very good. Therefore, by constructing a ranked list of terms similar to a user's keywords, we can retrieve images capturing correct activities.

2 RELATED SYSTEMS AND RESEARCH

Although lifelog retrieval is a relatively recent research topic, there have been a number of notable systems developed over the past decade. In this review, we focus on relevant systems that have been designed to support a user in searching through a lifelog archive, as opposed to simply browsing along some facets, such as date/time. In 2016, the *LeMoRe* system [2] used visual semantic concepts from lifelog images as the main indexing source. Along with this they supported the user by enhancing the annotations using *WordNet*. Arising from the previous instance of the Lifelog Search Challenge [7], we note six different retrieval systems which have taken part. Three of these systems performed well and we now highlight each of these. The winning system developed a novel approach for lifelog exploration using virtual reality system (*HTC VIVE*) with different methods of user interaction to increase the speed of searching for relevant moments: distance-based and contact-based interaction [4]. This system illustrated that a well-considered retrieval interface operating over the provided concept annotations would perform as well as, or better than, more sophisticated systems, such as the following: The lifelog interactive retrieval systems *LifeXplore* [11] converted a sequence of temporal lifelog images into videos, and utilized a 3-dimensional feature map to show all keyframes and retrieval results as well as implementing many complex features from video search; Additionally, the *VIRET* system [10] aimed to

LSC'19, June 10–13, 2019, Ottawa, ON, Canada

ACM ISBN 978-1-4503-6781-3/19/06…$15.00
https://doi.org/10.1145/3326460.3329161

evaluate the performance of a purely content-based video retrieval tool for Lifelog data. The *VIRET* system enhanced the provided visual concept annotations using a retrained GoogleNet (with an own set of 1,390 ImageNet labels) and supported user sketch input. Findings from these two systems suggested that the integration of additional Lifelog modalities would be an important for effective filtering.

The system presented in this paper builds on-top of the learnings from these early interactive lifelog retrieval systems. We focused on the user interacting challenges, so we utilised the provided visual concepts, but we extended these using the BERT algorithm to augment the provided annotations with a semantically richer and broader set of indexable terms for every image in the collection. We also designed the system interface to support fast and effective querying by non-expert users.

3 LIFELOG DATA

The dataset (described in [8]) contains images collected at the frequency of one every 30 seconds for a duration of 27 days from a lifelogging camera worn by one person. The photos are taken automatically by a camera so there is no focus on any particular object or purpose behind the capture. In addition, the dataset contains metadata consisting of:

- daily information such as number of calories, number of steps, sleep duration etc.
- health log (cholesterol, mood etc), food log, drink log
- per-minute data: location, image, bodymetric etc.

Note that images come with an automatic labeling of concepts with scores, so there is a set of indexable content provided.

4 PROPOSED APPROACH

4.1 System Design

The system has two components: (i) an offline component that processes the metadata and images utilizing natural language processing tool to create indexable content for the search engine, (ii) an online component for a user to interact with the search engine, to provide criteria, keywords and questions as well as to filter query results.

4.2 Semantic Preprocessing

In lifelogging data, concept labels are generated automatically by some API, so, they can be too general (for instance, instead of labeling an image "pizza", API just assigns a general term "food"). Additionally, in some cases, API's labels are just not accurate enough, most probably because the labeling tools have been trained on conventional photos and not lifelog photos. Therefore, in this paper, we utilize *BERT* [3], a state-of-the-art technique in natural language processing, to create ranked lists of "nearby" terms for a number of concepts (location, food, visual concepts etc.). *BERT* provides a novel embedding method for words, phrases and sentences in which the output vectors capture the semantics of natural language to some extent. For example, the embedding vector for the word "cake" is closer (in cosine similarity) to the embedding vector for the word "food" compared to the embedding vector of the word "dog". When receiving a user's request to search for images, each concept

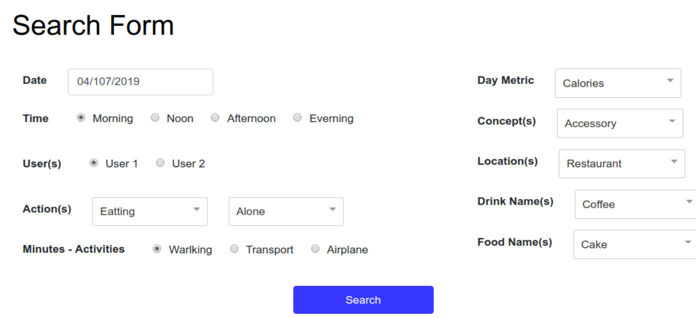

Figure 1: Search form

in the ranked lists corresponding to the user's provided keywords will be combined into queries to find all matching images. Finally, the results will be grouped and displayed for the user to select the right images.

Semantic Preprocessing and Enrichment is performed according to the following approach:

(1) Each concept in the metadata is embedded into a 768-dimensional vector space using bidirectional encoder representations from transformers (BERT). The embedding mechanism ensures that two concepts similar or "close" in meanings generate embedding vectors with short distances. To be more precise, the distance measure is the cosine similarity distance.
(2) Numeric vectors allow us to create a ranked list of terms for each concept provided (smallest cosine similarity distances to keywords). The first five terms in each list will be used to query the database.
(3) A dictionary of ranked lists is built and saved to the database for usage in the searching step.

To sum up, in our database, there are two components: (i) a hierarchical database consisting of meta data and images collected by two users, (ii) a dictionary where each concept (location, food, visual concept etc.) has to a list of similar terms ranked by the algorithm BERT.

4.3 User Interaction

We designed a simple user interface with a faceted search form as in Figure 1. The form allows user to enter different kinds of information including date, time of the date (Morning/ Noon/ Afternoon/ Evening), general day metric (calories etc.), location, action and so on.

During a search session, a user can choose to add more or to update his/her selections in the search form so that the search engine produces refined results. The search engines send all keywords collected from the search form to the internal search unit which looks up the dictionary (constructed in section 4.2) for similar terms to extend the list of keywords. Next, the search unit uses the list of keywords to query the database of meta data and images to extract images. The results are then displayed with sliders which let the user scroll left and right to pick relevant images to submit (see

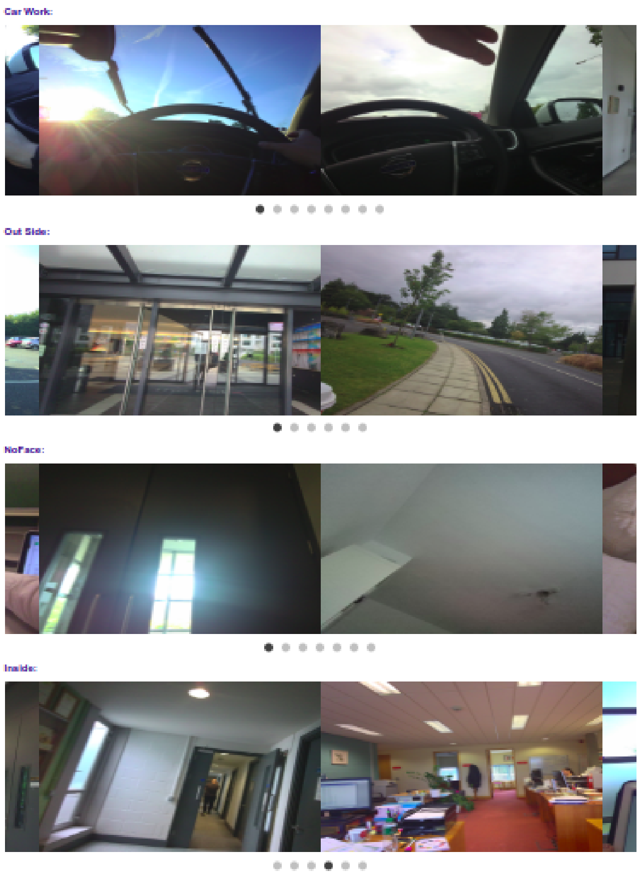

Figure 2: Results from our Search Engine

Figure 2. The first slider shows images that correspond to exact keywords provided by users. The second slider are for images queried by the closest terms of the provided keywords. The third slider uses the second closest terms to the keywords and so on.

A user should first check images in Slider 1 to see whether correct images are listed. Then, she/he moves down to Slider 2, 3 and 4 to find images. Slider 1 usually contains the results for the search, just in case the annotation is not accurate, Slider 2,3 and 4 will be useful in finding right images.

Selected images can either be immediately submitted for evaluation (at the LSC) or otherwise they can be stored in a temporary archive (using a shopping-cart metaphor) for later review.

5 CONCLUSION

In this paper, we described a first-generation prototype interactive lifelog retrieval system called *VieLens*. The system was developed to allow for the indexing and retrieval of lifelog content based on the ADLs of humans in everyday life. The key new idea is the use of the algorithm BERT in natural language processing to generate a dictionary of closest terms for each concept. The dictionary then helps extend the set of keywords to be inserted into queries to find images. Finally, a user can check and choose the right results interactively on a website.

6 ACKNOWLEDGMENTS

This paper is written under the VIBE 2018 funding from the Irish Embassy in Vietnam. Besides, we would also like to thank the Adapt Center under SFI Grant number 13/RC/2106 and the Insight Centre under SFI Grant number SFI/12/RC/2289. We also thank the University of Economics and Law, VNU-HCM, Vietnam for their generous support throughout the project.

REFERENCES

[1] Junho Choi, Chang Choi, Hoon Ko, and Pankoo Kim. 2016. Intelligent Healthcare Service Using Health Lifelog Analysis. *Journal of Medical Systems* 40 (2016), 1–10.

[2] Gabriel de Oliveira Barra, Alejandro Cartas Ayala, Marc Bolaños, Mariella Dimiccoli, Xavier Giró i Nieto, and Petia Radeva. 2016. 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.*

[3] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805* (2018).

[4] Aaron Duane, Cathal Gurrin, and Wolfgang Huerst. 2018. Virtual Reality Lifelog Explorer: Lifelog Search Challenge at ACM ICMR 2018. In *Proceedings of the 2018 ACM Workshop on The Lifelog Search Challenge (LSC '18).* ACM, New York, NY, USA, 20–23. https://doi.org/10.1145/3210539.3210544

[5] L Dubourg, A R Silva, C Fitamen, C J A Moulin, and C Souchay. 2016. SenseCam: A new tool for memory rehabilitation? *Revue Neurologique* 172, 12 (2016), 735–747. https://doi.org/10.1016/j.neurol.2016.03.009

[6] Jim Gemmell, Chester Bell, and Roger Lueder. 2006. MyLifeBits: A personal database for everything. *Commun. ACM* 49 (01 2006), 89–95. https://doi.org/10.1145/1107458.1107460

[7] Cathal Gurrin, Klaus Schoeffmann, Hideo Joho, Andreas Leibetseder, Liting Zhou, Aaron Duane, Duc-Tien Dang-Nguyen, Michael Riegler, Luca Piras, Minh-Triet Tran, Jakub Lokoč, and Wolfgang Hürst. 2019. Comparing Approaches to Interactive Lifelog Search at the Lifelog Search Challenge (LSC2018). *ITE Transactions on Media Technology and Applications* 7, 2 (2019), 46–59. https://doi.org/10.3169/mta.7.46

[8] Cathal Gurrin, Klaus Schoeffmann, Hideo Joho, Bernd Münzer, Rami Albatal, Frank Hopfgartner, Liting Zhou, and Duc-Tien Dang-Nguyen. 2019. A Test Collection for Interactive Lifelog Retrieval. In *MultiMedia Modeling - 25th International Conference, MMM 2019, Thessaloniki, Greece, January 8-11, 2019, Proceedings, Part I.* 312–324. https://doi.org/10.1007/978-3-030-05710-7_26

[9] Yongjin Kwon, Kyuchang Kang, Changseok Bae, Hee-Joon Chung, and Ju Han Kim. 2014. Lifelog Agent for Human Activity Pattern Analysis on Health Avatar Platform. In *Healthcare informatics research.*

[10] Jakub Lokoč, Tomáš Souček, and Gregor Kovalčik. 2018. Using an Interactive Video Retrieval Tool for LifeLog Data. In *Proceedings of the 2018 ACM Workshop on The Lifelog Search Challenge (LSC '18).* ACM, New York, NY, USA, 15–19. https://doi.org/10.1145/3210539.3210543

[11] Bernd Münzer, Andreas Leibetseder, Sabrina Kletz, Manfred Jürgen Primus, and Klaus Schoeffmann. 2018. lifeXplore at the Lifelog Search Challenge 2018. In *Proceedings of the 2018 ACM Workshop on The Lifelog Search Challenge (LSC '18).* ACM, New York, NY, USA, 3–8. https://doi.org/10.1145/3210539.3210541

[12] L N Signal, J Stanley, M Smith, M B Barr, T J Chambers, J Zhou, A Duane, C Gurrin, A F Smeaton, C McKerchar, A L Pearson, J Hoek, G L S Jenkin, and C Ni Mhurchu. 2017. Children's everyday exposure to food marketing: an objective analysis using wearable cameras. *International Journal of Behavioral Nutrition and Physical Activity* 14, 1 (oct 2017), 137. https://doi.org/10.1186/s12966-017-0570-3