Exquisitor at the Lifelog Search Challenge 2020

Omar Shahbaz Khan
IT University of Copenhagen
Copenhagen, Denmark
omsh@itu.dk

Mathias Dybkjær Larsen
IT University of Copenhagen
Copenhagen, Denmark
mdyl@itu.dk

Liam Alex Sonto Poulsen
IT University of Copenhagen
Copenhagen, Denmark
liap@itu.dk

Björn Þór Jónsson
IT University of Copenhagen
Copenhagen, Denmark
bjorn@itu.dk

Jan Zahálka
Czech Technical University in Prague
Prague, Czech Republic
jan.zahalka@cvut.cz

Stevan Rudinac
University of Amsterdam
Amsterdam, Netherlands
s.rudinac@uva.nl

Dennis Koelma
University of Amsterdam
Amsterdam, Netherlands
d.c.koelma@uva.nl

Marcel Worring
University of Amsterdam
Amsterdam, Netherlands
m.worring@uva.nl

ABSTRACT

We present an enhanced version of Exquisitor, our interactive and scalable media exploration system. At its core, Exquisitor is an interactive learning system using relevance feedback on media items to build a model of the users' information need. Relying on efficient media representation and indexing, it facilitates real-time user interaction. The new features for the Lifelog Search Challenge 2020 include support for timeline browsing, search functionality for finding positive examples, and significant interface improvements. Participation in the Lifelog Search Challenge allows us to compare our paradigm, relying predominantly on interactive learning, with more traditional search-based multimedia retrieval systems.

CCS CONCEPTS

• **Information systems → Multimedia and multimodal retrieval**; **Multimedia databases**.

KEYWORDS

Lifelogging; Interactive learning; Exquisitor.

ACM Reference Format:
Omar Shahbaz Khan, Mathias Dybkjær Larsen, Liam Alex Sonto Poulsen, Björn Þór Jónsson, Jan Zahálka, Stevan Rudinac, Dennis Koelma, and Marcel Worring. 2020. Exquisitor at the Lifelog Search Challenge 2020. In *Proceedings of the Third Annual Workshop on the Lifelog Search Challenge (LSC '20), June 9, 2020, Dublin, Ireland*. ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/3379172.3391718

1 INTRODUCTION

The Lifelog Search Challenge (LSC) is a live system-evaluation event, where researchers compare their systems based on their ability to help users quickly solve search-related tasks for a multimodal lifelog dataset. Each task in LSC is an independent query, to be solved in a few minutes, where a correct result is a single image returned from a set of relevant images. The query description is given gradually, as might be typical when a lifelog is used to find information and the user slowly remembers more details about the situation. The first two editions of LSC, held in 2018 [3, 4] and 2019 [5], have showcased a variety of multimedia retrieval systems aiming to search the lifelog with different approaches, ranging from traditional keyword search to novel virtual reality-based approaches (e.g., see [1, 9, 10, 12]).

We have recently developed Exquisitor, a highly scalable interactive learning system for general multimedia analytics applications [7]. When applied to LSC, the user is initially presented with a set of randomly selected images from the lifelog and asked to give feedback on (some of) the items about their relevance to the LSC task at hand. The feedback is used to build (and subsequently update) a classification model, which in turn is used to provide new suggestions; this iterative process continues as long as the user deems necessary. Figure 1 describes Exquisitor's interactive learning interface. A key feature that sets Exquisitor apart from other interactive learning approaches is its scalability: Exquisitor can retrieve suggestions from the LSC 2020 collection of 43K images in less than 50 milliseconds using a single CPU core, allowing to retrieve suggestions very rapidly following each user interaction.

Exquisitor participated in LSC 2019 [8], where it ranked sixth out of nine participants. The main lesson from LSC 2019 was that interactive learning is a viable approach, even in this heavily search-oriented competition setting. However, we also identified some shortcomings of the Exquisitor system itself that prevented solving some of the tasks. In this paper, we present the lessons learned from LSC 2019 and how we have improved the system for participation in LSC 2020. These improvements were partly implemented for participation in the Video Browser Showdown 2020 [6], where Exquisitor ranked fifth out of eleven participants.

LSC '20, June 9, 2020, Dublin, Ireland

ACM ISBN 978-1-4503-7136-0/20/06…$15.00
https://doi.org/10.1145/3379172.3391718

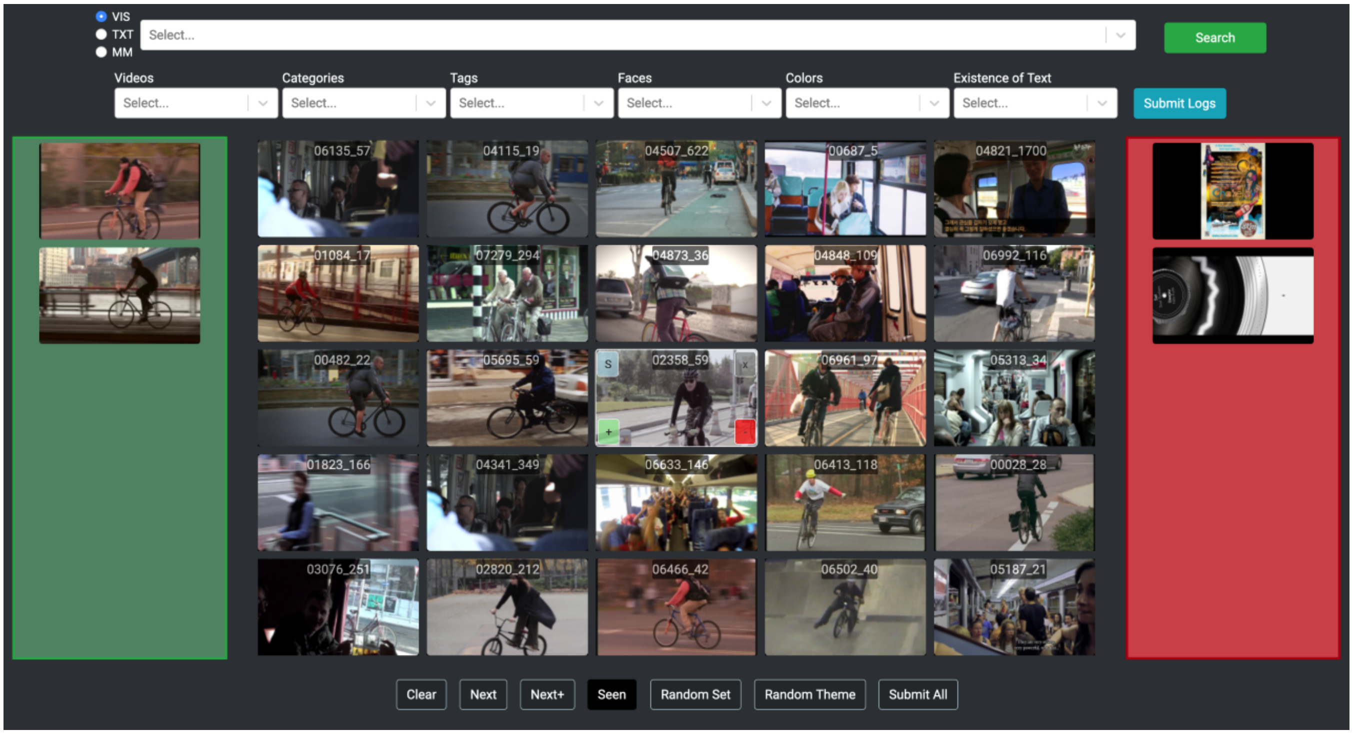


Figure 1: Exquisitor's interactive learning interface. Previously selected positive examples are shown on the left and negative examples on the right. The middle panel shows 25 suggestions based on the classification model built based on the user's feedback. By hovering over a thumbnail (see the middle thumbnail), users can select the image/video clip as a positive or negative example (bottom left/right corners), remove it from consideration (upper right corner) or submit as solution to the task (upper left corner). The top bars are for search and filtering, as described in the text.

The remainder of the paper is organized as follows. Section 2 briefly outlines the Exquisitor approach. Section 3 describes the lessons learned from participation in LSC 2019, and Section 4 reviews the changes made to Exquisitor based on those lessons.

2 EXQUISITOR

Exquisitor is a state-of-the-art multimodal interactive learning approach that combines efficient representation of data, a fast interactive classifier, and large-scale collection indexing [7]. The data representation for each multimodal item comprises state-of-the-art semantic visual concepts and text features. The semantic features are compressed per modality using an index-based compression method [16] that achieves over 99% compression rate whilst yielding a data representation that preserves the semantic information in the original data. The interactive classifier of choice, linear SVM, operates directly in the compressed space to greatly speed up the suggestion retrieval process. While more complex models, such as those based on CNN architectures, have achieved great successes in supervised learning settings, the performance of linear models for classification is still unparalleled in interactive learning due to their relatively good performance, explainability and the ability to scale to very large collections [7, 11, 13, 16].

To build an index suitable of scaling up to large scale datasets, Exquisitor builds on the extended Cluster Pruning (eCP) algorithm [2], which creates a hierarchical structure of the collection and enables efficient weaving of index utilization into the interactive learning pipeline. Instead of scoring all items in the collection with the classifier trained on user input, in each interaction round, Exquisitor first identifies the b clusters most relevant to the query, based on the SVM model, and then only scores items in those clusters, again using the SVM model to produce the suggestion candidates per modality. More specifically, the b clusters of each modality are divided into s segments, and a list of r candidates is produced from each segment. The final suggestions are then obtained by performing late modality fusion over the $s \times r$ candidates from each modality to produce the final k suggestions for the user.

By using a high-dimensional index, Exquisitor's suggestion retrieval relies not only on the scores provided by the interactive classifier, but also harnesses the collection's high-dimensional structure; our results indicate that this can indeed improve the quality of the suggestions at scale. In [7], large-scale, artificial actor-simulated experiments [15] with the ImageNet and YFCC100M collections show that with parameter settings of $b = 256, s = 16, r = 1,000$ and $k = 25$, Exquisitor significantly outperforms the state of the art in user relevance feedback.

3 LESSONS FROM LSC 2019

As outlined in the introduction, we believe that interactive learning as a concept performed quite well on the search-based tasks of LSC 2019. We found, however, that the system was missing some features that would have been useful for solving some of the tasks:

- *Model Bootstrapping.* Initially, the user is presented with a screen of 25 random images from the lifelog collection. Even for the relatively small LSC 2019 collection of about 43K images, this represents less than 0.1% of the collection. For some tasks there were few positive examples in the collection, so the odds of randomly finding positive examples was therefore very low. Some means of searching for positive examples is thus clearly needed.
- *Temporal Overview.* Several LSC tasks described a sequence of events leading up to the correct answer to the task, and sometimes these prior events were easier to identify than the eventual answer. Without any means to browse a timeline, finding these prior events offered limited value for solving the tasks.
- *General Interface Issues.* We found that the interactive learning interface itself had multiple problems, and was in particular difficult to use for novice users. This included basic issues such as too much unused space on the screen and too many mouse-clicks for common operations, as well as requiring complex interactions to apply filters to the relevance feedback process.
- *Metadata Integration.* Finally, at LSC 2019 we used only a subset of the available metadata. While the subset we used would have been sufficient to solve most of the tasks, integrating all available metadata is important for the ability to solve general analytics tasks.

We believe that these findings apply generally for any multimedia analytics application, as the problems encountered during LSC could be encountered in many situations where a combination of search and exploration is required.

4 NEW FEATURES FOR LSC 2020

In order to address the lessons described above, we have implemented the following changes to the Exquisitor system:

- *Model Bootstrapping.* We have implemented text-search functionality, using pylucene, over the metadata of the lifelog images, including the semantic concepts and their descriptions. Note, however, that the primary goal of the search functionality is not to find the answers to the tasks—although this may happen in some cases—but rather to identify positive example images, or even specific negative example images, that can be used to build the model of user intent.
- *Temporal Overview.* For the Video Browser Showdown, we implemented a video explorer to browse short scenes within the context of the videos, as shown in Figure 2. By considering each lifelog image as a thumbnail from a video (albeit, a video with a very low frame-rate), we adapt this functionality to support timeline browsing within the lifelog collection. We have also improved the timeline explorer implementation to provide flexible granularity of the lifelog timeline, thus providing better overview for the user.
- *General Interface Issues.* In order to improve usability, we have eliminated some functionality that was not used in practice (e.g., incrementally replacing images with new suggestions), streamlined several important operations (e.g., examining the collections of positive or negative examples), and improved screen usage significantly by eliminating unused background space.
- *Metadata Integration.* Finally, we are working to improve the use of images and metadata. We have applied state-of-the-art ResNeXt-101 visual concept detectors [14] to the lifelog images, impacting both the user relevance feedback process and text search. We have also improved the filtering process and are working to extend the range of metadata from the collection that is available to users. As an example, the ability to filter lifelog images based on geo-location could potentially be important for some LSC tasks.

As noted above, some of these enhancements have already been applied in our participation in the Video Browser Showdown 2020. With the additional changes made for LSC participation, we expect that the system will perform significantly better with LSC tasks.

5 CONCLUSION

Exquisitor is an efficient interactive learning system, which relies on user relevance feedback to build a model of the user's information need. While Exquisitor targets general multimedia analytics applications, the participation in the Lifelog Search Challenge (LSC) nevertheless allows comparison with more traditional search-based media retrieval systems. In this paper we have described the lessons learned from participation in LSC 2019 and the changes made to the Exquisitor system for our participation in LSC 2020.

ACKNOWLEDGMENTS

This work was supported by a PhD grant from the IT University of Copenhagen and by the European Regional Development Fund (project Robotics for Industry 4.0, CZ.02.1.01/0.0/0.0/15 003/0000470).

REFERENCES

[1] Aaron Duane, Cathal Gurrin, and Wolfgang Hürst. 2018. Virtual Reality Lifelog Explorer: Lifelog Search Challenge at ACM ICMR 2018. In *Proceedings of the ACM Workshop on Lifelog Search Challenge, LSC 2019.* ACM, Yokohama, Japan, 20–23.

[2] Gylfi Þór Guðmundsson, Björn Þór Jónsson, and Laurent Amsaleg. 2010. A Large-scale Performance Study of Cluster-based High-dimensional Indexing. In *Proc. International Workshop on Very-large-scale Multimedia Corpus, Mining and Retrieval (VLS-MCMR).* ACM, Firenze, Italy, 31–36.

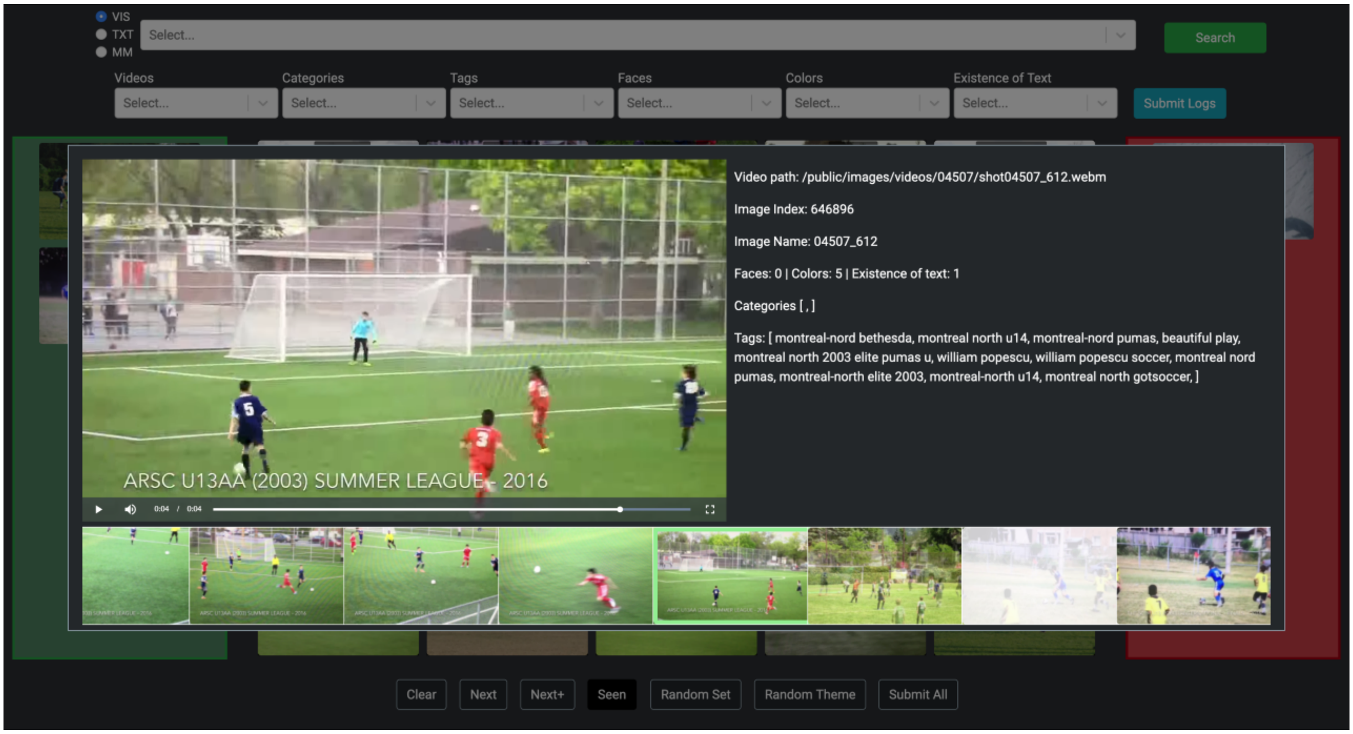
[3] Cathal Gurrin, Klaus Schoeffmann, Hideo Joho, Duc-Tien Dang-Nguyen, Michael Riegler, and Luca Piras (Eds.). 2018. *Proceedings of the ACM Workshop on Lifelog Search Challenge, LSC 2018.* ACM, Yokohama, Japan.

[4] Cathal Gurrin, Klaus Schoeffmann, Hideo Joho, Andreas Leibetseder, Liting Zhou, Aaron Duane, Duc-Tien Dang-Nguyen, Michael Riegler, Luca Piras, Minh-Triet Tran, et al. 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.

[5] Cathal Gurrin, Klaus Schöffmann, Hideo Joho, Duc-Tien Dang-Nguyen, Michael Riegler, and Luca Piras (Eds.). 2019. *Proceedings of the ACM Workshop on Lifelog Search Challenge, LSC 2019.* ACM, Ottawa, ON, Canada.

[6] Björn Þór Jónsson, Omar Shahbaz Khan, Dennis C. Koelma, Stevan Rudinac, Marcel Worring, and Jan Zahálka. 2020. Exquisitor at the Video Browser Showdown 2020. In *Proceedings of the International Conference on MultiMedia Modeling (MMM).* Springer, Daejeon, South Korea, 796–802.

[7] Omar Shahbaz Khan, Björn Þór Jónsson, Stevan Rudinac, Jan Zahálka, Hanna Ragnarsdóttir, Þórhildur Þorleiksdóttir, Gylfi Þór Guðmundsson, Laurent Amsaleg, and Marcel Worring. 2020. Interactive Learning for Multimedia at Large. In



Video path: /public/images/videos/04507/shot04507_612.webm

Image Index: 646896

Image Name: 04507_612

Faces: 0 | Colors: 5 | Existence of text: 1

Categories [,]

Tags: [montreal-nord bethesda, montreal north u14, montreal-nord pumas, beautiful play, montreal north 2003 elite pumas u, william popescu, william popescu soccer, montreal nord pumas, montreal-north elite 2003, montreal-north u14, montreal north gotsoccer,]

Figure 2: Exquisitor's interface for exploring media items (images or video clips) in a temporal context. The interface shows details of the metadata associated with the media item, and allows exploration of the temporal context.

Proceedings of the European Conference on Information Retrieval (ECIR). Springer, Lisboa, Portugal, 16.

[8] Omar Shahbaz Khan, Björn Þór Jónsson, Jan Zahálka, Stevan Rudinac, and Marcel Worring. 2019. Exquisitor at the Lifelog Search Challenge 2019. In *Proceedings of the ACM Workshop on Lifelog Search Challenge, LSC 2019.* ACM, Ottawa, ON, Canada, 7–11.

[9] Andreas Leibetseder, Bernd Münzer, Manfred Jürgen Primus, Sabrina Kletz, Klaus Schoeffmann, Fabian Berns, and Christian Beecks. 2019. lifeXplore at the Lifelog Search Challenge 2019. In *Proceedings of the ACM Workshop on Lifelog Search Challenge, LSC 2019.* ACM, Ottawa, ON, Canada, 13–17.

[10] Jakub Lokoč, Gregor Kovalčík, Tomáš Souček, Jaroslav Moravec, and Přemysl Čech. 2019. VIRET: A Video Retrieval Tool for Interactive Known-Item Search. In *Proc. ACM International Conference on Multimedia Retrieval (ICMR).* ACM, Ottawa, ON, Canada, 177–181.

[11] Ionuţ Mironică, Bogdan Ionescu, Jasper Uijlings, and Nicu Sebe. 2016. Fisher Kernel Temporal Variation-based Relevance Feedback for video retrieval. *Computer Vision and Image Understanding* 143 (2016), 38 – 51. https://doi.org/10.1016/j.cviu.2015.10.005 Inference and Learning of Graphical Models Theory and Applications in Computer Vision and Image Analysis.

[12] Luca Rossetto, Ralph Gasser, Silvan Heller, Mahnaz Amiri Parian, and Heiko Schuldt. 2019. Retrieval of Structured and Unstructured Data with vitrivr. In *Proceedings of the ACM Workshop on Lifelog Search Challenge, LSC 2019.* ACM, Ottawa, ON, Canada, 27–31.

[13] Sudheendra Vijayanarasimhan, Prateek Jain, and Kristen Grauman. 2014. Hashing Hyperplane Queries to Near Points with Applications to Large-Scale Active Learning. *IEEE Trans. Pattern Anal. Mach. Intell.* 36, 2 (2014), 276–288.

[14] Saining Xie, Ross B. Girshick, Piotr Dollár, Zhuowen Tu, and Kaiming He. 2017. Aggregated Residual Transformations for Deep Neural Networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR).* IEEE Computer Society, Honolulu, HI, USA, 5987–5995.

[15] Jan Zahálka, Stevan Rudinac, and Marcel Worring. 2015. Analytic Quality: Evaluation of Performance and Insight in Multimedia Collection Analysis. In *Proc. ACM Multimedia.* ACM, Brisbane, Australia, 231–240.

[16] Jan Zahálka, Stevan Rudinac, Björn Þór Jónsson, Dennis C. Koelma, and Marcel Worring. 2018. Blackthorn: Large-Scale Interactive Multimodal Learning. *IEEE Transactions on Multimedia* 20, 3 (2018), 687–698.