Exquisitor at the Lifelog Search Challenge 2021: Relationships Between Semantic Classifiers

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

Aaron Duane
IT University of Copenhagen
Copenhagen, Denmark
aadu@itu.dk

Björn Þór Jónsson
IT University of Copenhagen
Copenhagen, Denmark
bjth@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

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

ABSTRACT

Exquisitor is a scalable media exploration system based on interactive learning. To satisfy a user's information need, the system asks the user for feedback on media items and uses that feedback to interactively construct a classifier, that is in turn used to identify the next potentially relevant set of media items. To facilitate effective exploration of a collection, the system offers filters to narrow the scope of exploration, search functionality for finding good examples for the classifier, and support for timeline browsing of videos or image sequences. For this year's Lifelog Search Challenge, we have enhanced Exquisitor to better support tasks with a temporal component, by adding features that allow the user to build multiple classifiers and merge the classifier results, using both traditional set operators and advanced temporal operators.

CCS CONCEPTS

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

KEYWORDS

Lifelogging; Interactive learning; Exquisitor; Temporal Relations

ACM Reference Format:
Omar Shahbaz Khan, Aaron Duane, Björn Þór Jónsson, Jan Zahálka, Stevan Rudinac, and Marcel Worring. 2021. Exquisitor at the Lifelog Search Challenge 2021: Relationships Between Semantic Classifiers. In *Proceedings of the 4th Annual Lifelog Search Challenge (LSC '21), August 21, 2021, Taipei, Taiwan.* ACM, New York, NY, USA, 4 pages. https://doi.org/10.1145/3463948.3469255

1 INTRODUCTION

Exquisitor is a highly scalable interactive learning system for general multimedia analytics applications [4]. At its core, it is a user relevance feedback system where the user provides positive or negative feedback on a set of images, which is used to build or update its classifier, that in turn is used to select new suggestions for the user to give feedback on. This iterative process continues as long as the user deems necessary. Aside from this relevance feedback loop, Exquisitor also provides users with metadata filters to narrow the scope of exploration, a keyword search feature to more quickly find positive items for its classifier, and a timeline to explore videos or sequences of images [6]. Exquisitor's scalability for interactive learning has been proven capable of achieving sub-second latency for collections with 100M images [3].

The Lifelog Search Challenge (LSC) is an annual interactive event, now in its fourth year, where multiple multimedia retrieval systems and their users compete in solving search-related tasks for a multimodal lifelog dataset. A task in LSC is formulated with a textual description of the information need, which is revealed gradually over a time period of a few minutes, with each step adding or clarifying details. For each task there is a relevant set of one or more images and the objective for participants is to find one image within that set. This measure of providing information gradually reflects a user recalling an event from memory, where over time more details about and around the event become clearer. Once the user finds one image of the desired event the rest can surely be found through that. Participants' scores are determined mostly by how fast they can solve tasks, but they are also penalized if they submit incorrect images, so there is an incentive to be both fast and correct. LSC 2021 uses a collection with roughly 183K images and corresponding metadata for four month-long periods from a lifelog [2].

Exquisitor participated in both LSC 2019 [5] and LSC 2020 [6], where it ranked 6th out of 9 and 6th out of 14 teams respectively. The main positive lesson from LSC 2020 was that the features implemented to augment the relevance feedback process to train an interactive classifier, such as searching for positive examples and navigating the temporal structure of the lifelog collection, work well for solving most tasks. While Exquisitor was the only team able to solve all tasks with a temporal component in LSC 2020, the systems that can specify temporal relations between multiple queries, generally solved these tasks faster. We therefore believe that the time to solve temporal tasks can be significantly improved.

In this paper, we therefore describe new features that allow specifying temporal relations between the results of classifiers.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.
LSC '21, August 21, 2021, Taipei, Taiwan
© 2021 Association for Computing Machinery.
ACM ISBN 978-1-4503-8533-6/21/08…$15.00
https://doi.org/10.1145/3463948.3469255

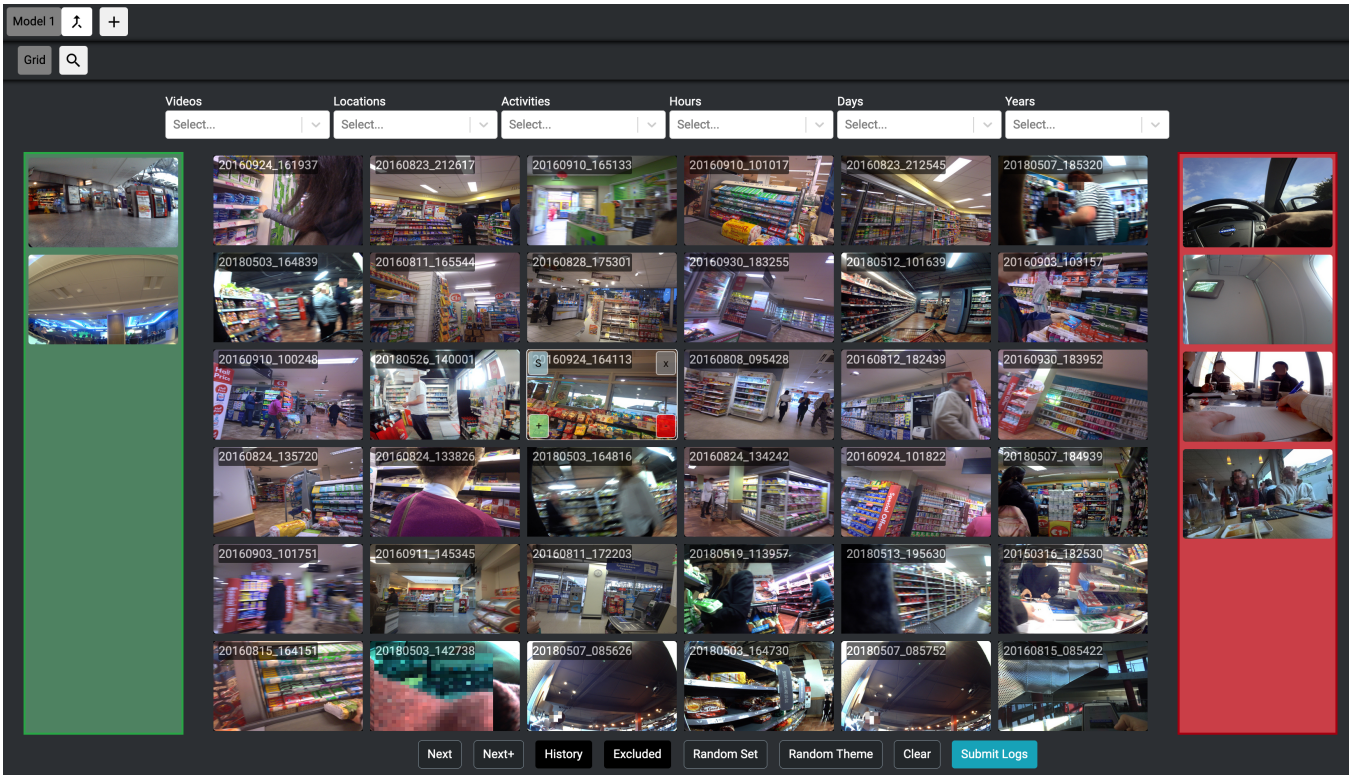


Figure 1: Exquisitor's interactive learning interface, after several feedback rounds. Previously selected positive examples are shown in the green container on the left and negative examples in the red container on the right. The middle section shows 36 suggestions based on the current classifier, which was built using the positive and negative examples. 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 the thumbnail as a solution to the competition task (upper left corner). The bars above the suggested images are for applying filters on metadata. The search icon button above the filters opens keyword-search tabs that can be used to find examples.

The user can start multiple relevance feedback processes to train classifiers for specific concepts. These can be merged together with temporal constraints to define a temporal relation between them. Using an example task, we demonstrate that this can lead to faster completion of some tasks with a temporal relationship.

The remainder of the paper is organized as follows. Section 2 briefly outlines the Exquisitor system and features from LSC 2020, while Section 3 presents the changes made to Exquisitor for LSC 2021. We conclude the paper in Section 4.

2 EXQUISITOR

Exquisitor is a state-of-the-art large-scale multimodal interactive learning system, that works with an efficient compressed data representation of semantic multimodal features, a fast interactive classifier and a scalable high-dimensional index [4]. The semantic features are compressed using a method [9] that achieves over 99% compression rate whilst preserving the top semantic features' information from the original data. The interactive classifier is a linear SVM, with an unparalleled interactive learning performance in terms of explainability and scalability [4, 7–9]. The high-dimensional index of Exquisitor is based on the extended Cluster Pruning (eCP) algorithm [1] which prioritizes latency over precision by attempting to form similar size clusters.

Figure 1 shows the interface used for relevance feedback in Exquisitor. The relevance feedback process starts by the user providing positive or negative feedback to an arbitrary suggestion set. Once an item is deemed positive or negative, the classifier is trained and used to retrieve new suggestions. To retrieve suggestions, the b clusters most likely to contain candidates are selected using the index. These clusters are then divided into s segments, each of which is processed to return k suggestions. If multiple modalities are used, rank aggregation is used at this stage to combine them. From the combined sk suggestions from the segments, the top k suggestions overall are selected and presented to the user for another round of feedback. This process continues until the user is satisfied.

Aside from Exquisitor's core component of relevance feedback, the system also provides the following features to support the learning process:

Metadata Filters The system is augmented with filters on metadata to delimit the scope of the exploration. For the

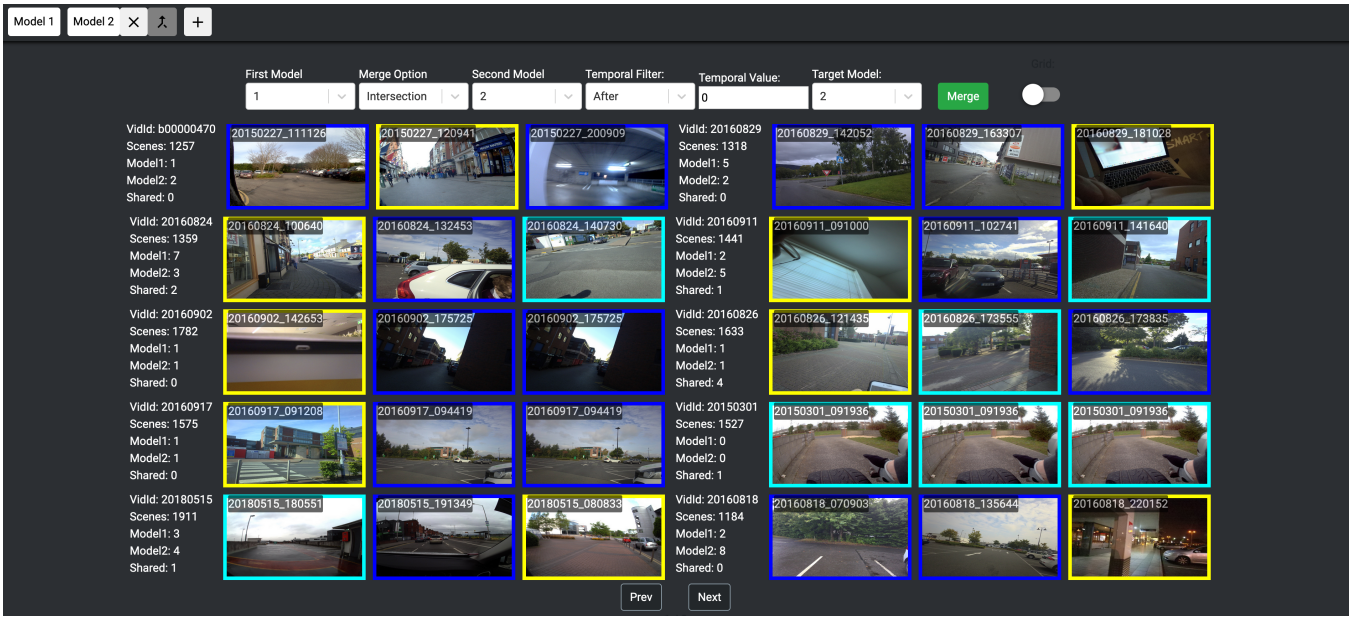


Figure 2: The merge view for two classifiers. Where the first classifier aims to find footpaths and the second to find parking lots. The merge view intersects the two classifiers suggestion sets. The light-blue highlighted images are suggestions from both classifiers while the yellow are from classifier 1 and blue from classifier 2.

LSC collection, filters are included on location, activities, hours of the day, days of the week, and the year. Filters are applied to media items before selecting the top k items from each segment.

Keyword Search A common problem in interactive learning is finding the first positive examples to start training the classifier. We observed this clearly in LSC 2019, where the arbitrary starting points often resulted in many rounds of providing negatives only. By providing search functionality in LSC 2020, better positive examples could be found, leading to faster training and a shorter relevance feedback process.

Timeline Explorer Exquisitor supports a timeline browser adapted for LSC to explore days. Indeed, several tasks in each LSC competition describe multiple events happening around the main event that is desired, where timeline browsing is useful. Additionally Exquisitor has been equipped with an exclusion filter, which can be set on days that have been fully explored and deemed not to be relevant, and can thus be prevented from appearing again in the results.

Furthermore, there are many other less significant features, such as the possibility to review and re-evaluate the positive and negative examples, scroll through a historical view of already seen items, and obtain a random set of items for the next interaction round.

3 NEW FEATURES FOR LSC 2021

In some cases, creating a single classifier to satisfy an information need is difficult. When multiple different semantic concepts are desired, for example, one concept may dominate another in the collection, making it hard to balance the positive examples. And where those concepts relate to multiple separate events in a lifelog that occurred in sequence, building a single classifier to return all the events may be impossible. In this section we therefore describe two new features that have been added to the Exquisitor system for LSC 2021 to address such tasks.

Multiple Classifiers This feature enables Exquisitor to spawn any number of classifiers that allows Exquisitor to train for specific concepts. The suggestions from two classifiers can be merged through a merge view with a relationship operator op, where $op \in \{\cap, \cup, \setminus, -\}$. The results are ranked with the average score from the accumulated rank of the items from each classifier and the total number of scenes/images. They are displayed in an augmented timeline view which highlights items from each classifier present in the same video or sequence of images with set time intervals such as days.

Temporal Constraints When merging two classifiers additional options are available for applying temporal constraints. These constraints emphasize whether the result sets should prioritize items based on the order of classifiers, as in a classifier 2 item exists after a classifier 1 item within a given interval, or irregardless of order within a given time interval.

As an example of the use of these features, consider task "LSC27" from LSC 2019, which stated: "Walking on a green footpath, to my car. I remember I had come off a flight and it was around lunch-time. I got into my car and drove to have a meal. No, I drove to work where I had lunch." Footpaths and parking lots are in abundance in the LSC collection, but they do not appear together in any image, and the combination of the two in short sequence is very rare, if not unique.

Figure 2 illustrates how the new functionality can be used to solve the task. After training two classifiers, one for identifying footpaths and one for identifying parking lots, the figure shows the merging view for those two classifiers. This view shows the three most relevant images per day and highlights them based on whether they were in the result of classifier 1 (yellow), classifier 2 (blue), or both (light-blue). For this example query the relevant item was discovered among the suggestions of classifier 2 after an item from the merge view was set as positive for that classifier.

4 CONCLUSION

Exquisitor is a state-of-the-art large-scale interactive learning system which has participated in the two previous editions of the Lifelog Search Challenge. For this year's LSC we have added a feature to merge results of multiple classifiers, possibly with temporal constraints, to better identify complex events with temporal components.

ACKNOWLEDGMENTS

This work was supported by a PhD grant from the IT University of Copenhagen, the European Regional Development Fund (project Robotics for Industry 4.0, CZ.02.1.01/0.0/0.0/15 003/0000470), and by MCSA-IF grant 893914.

REFERENCES

[1] 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.

[2] Cathal Gurrin, Björn Þór Jónsson, Klaus Schöffmann, Duc-Tien Dang-Nguyen, Jakub Lokoč, Minh-Triet Tran, Wolfgang Hürst, Luca Rossetto, and Graham Healy. 2021. Introduction to the Fourth Annual Lifelog Search Challenge, LSC'21. In *Proc. International Conference on Multimedia Retrieval (ICMR'21).* ACM, Taipei, Taiwan.

[3] Björn Þór Jónsson, Omar Shahbaz Khan, Hanna Ragnarsdóttir, Þórhildur Þorleiksdóttir, Jan Zahálka, Stevan Rudinac, Gylfi Þór Guðmundsson, Laurent Amsaleg, and Marcel Worring. 2019. Exquisitor: Interactive Learning at Large. arXiv:1904.08689.

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

[5] 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.

[6] 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 Lifelog Search Challenge.* ACM.

[7] 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.

[8] 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.

[9] 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.