EARLY DRAFT: A COMPARATIVE REVIEW OF SYSTEM PERFORMANCE AT LSC'20

A PREPRINT

Naushad Alam*
Insight Research Centre
School of Computing
Dublin City University
Pittsburgh, PA 15213
hippo@cs.cranberry-lemon.edu

April 16, 2021

ABSTRACT

This paper provides analysis of the performance of systems participating in the third Lifelog Search Challenge (LSC'20) held at the 2020 Annual ACM International Conference on Multimedia Retrieval (ICMR).

***K*eywords** lifelog · information retrieval · multimodal · analytics

1 Introduction

Life-logging addresses the ambitious challenge of creating a digital record of the totality of an individual's experiences, captured through digital sensors and permanently stored as a multi-modal multimedia archive. Creation of such an extensive pool of data has a wide range of applications that will potentially improve the quality of life of individuals. The sheer size of datasets additionally creates a substantial challenge in terms of multi-modal data analytics and retrieval.

The LSC workshop has an associated dataset, a new four-month rich multimodal lifelog data set produced including images captured on a wearable-camera, user activity information as well as location.

LSC'20 had more than 150 attendees and was streamed live at the twitch URL[2]. As with all previous versions, LSC'20 was a highly interactive and entertaining workshop modelled on the successful Video Browser Showdown annual competition at the MMM conference. LSC is a participation workshop, which means that all participants will write and present a paper, as well as taking part in the live interactive search challenge.

LSC poses a unique multimodal information retrieval problem where participants are required to retrieve specific events/memories from a vast archive of multimedia data. LSC'20 evaluated the participating system on 24 topics, which were structured to mimic how humans incrementally recall information from their memory. The topics revealed information incrementally to participants (at 30s interval), starting at t=0s and going up to t= 150s, usually giving out the visual description of the scene/event early on followed by explicit details like time, date, location, etc. at later stages.For each topic, the teams were required to find and submit the ground truth image within the stipulated maximum time of 300s and were assigned scores based on a scoring mechanism, which also took into account the speed of submission (faster the better) as well as the count of incorrect submissions (lesser the better) before the correct submission was made.

The purpose of this paper is to provide analysis of system performance at LCS'20 to gain a better understanding of the factors that lead to high performance retrieval in the challenge. 14 teams participated in the competition last year and

*Use footnote for providing further information about author (webpage, alternative address)—*not* for acknowledging funding agencies.

[2] `https://www.twitch.tv/lscworkshop`

this paper provides analysis of what led to good performance of systems with respect to the strategy of the human user or searcher as well as the underlying systems.

2 Participating Systems

Table 2 gives a summary of the range of approaches taken by participating systems at LSC'20 and Table 3 includes performance results.

Add other precision results for systems to this single table and remove the individual ones.
priate

Below we provide analysis of performance of participating systems.

3 Performance Analysis

A total of 14 systems participated in the 2020 edition of Lifelog Search Challenge and were evaluated on 24 topics.

3.1 Individual Analysis

Figures 1, 2, 3, 4 and 5 provide a depiction of the speed by which individual systems retrieved the correct image for descriptive versus temporal topics.

Can we add the time limit to the caption above and below?

3.2 Comparative Analysis

This section aims to compare their performance of the 14 systems, that participated in the 2020 edition of Lifelog Search Challenge and were evaluated on 24 temporal or descriptive topics, in terms of speed and accuracy.

3.2.1 Correct - Incorrect Submissions

We firstly analyse the performance of participating systems based on the number of correct and incorrect submissions made across all evaluation topics. This yields a high level overview of the best-performing approaches. Figure 6 shows the distribution of correct and incorrect submission for each team. SOMHUNTER tops the chart with 20 correct submissions and systems like MYSCEAL, VIRET , LIFEXPLORE and VITRIVR closely follow suit with 19, 18, 18 and 17 correct submissions, respectively.

The precision of each team i.e. ratio of correct upon total submissions is also an important metric. High precision does not necessarily imply a high performing system but sheds light on the search strategy of teams and how restrained their approach was in submitting responses. Figure 16 shows the precision of all teams across all topics. VIRET and LifeGraph are two system with maximum precision among all teams but they are far apart in terms of their position on the final leaderboard.

3.2.2 Analysis of Time Distribution

Since time is a very crucial factor in the scoring mechanism for LSC, we analyze the distribution of time taken to correctly submit a response for every team across all topics. Figure 17 shows box and whisker plots of time distribution for correct responses across teams.Among the top-5 systems (left to right), MySceal has the lowest average time while SOMHunter has lowest median time to submit a correct response. From position 6 to 10 in the leaderboard, LifeSeeker is the fastest with lowest average and median time among others.

To further analyze the speed of competing systems we tried to observe the number of occurrences when the system was in top-3 to submit a correct response given a topic.Figure 18 shows count of occurrences to submit a correct response in top-3 plotted against every team. MySceal has the highest 11 occurrences in top-3 among the 19 topics it answered correctly. This is followed by VIRET and SOMHunter with 9 and 8 occurrences respectively.

3.2.3 Categorization of Topics - Analysis

LSC, 2020 evaluated the competing systems on 24 topics. Each of these topics follows a very similar structure where information is revealed to participants every 30 seconds starting from t=0s till t=150s. Usually, visual description of the scene/event is revealed early on followed by temporal events (if any) which is then followed by specific details like time or location, etc. However, some topics describe the scene/event in much greater detail with little emphasis on temporal events while others reveal much more information about temporal events as compared to scene descriptions.

Basis this distinction we segregated the evaluation topics into following two categories:

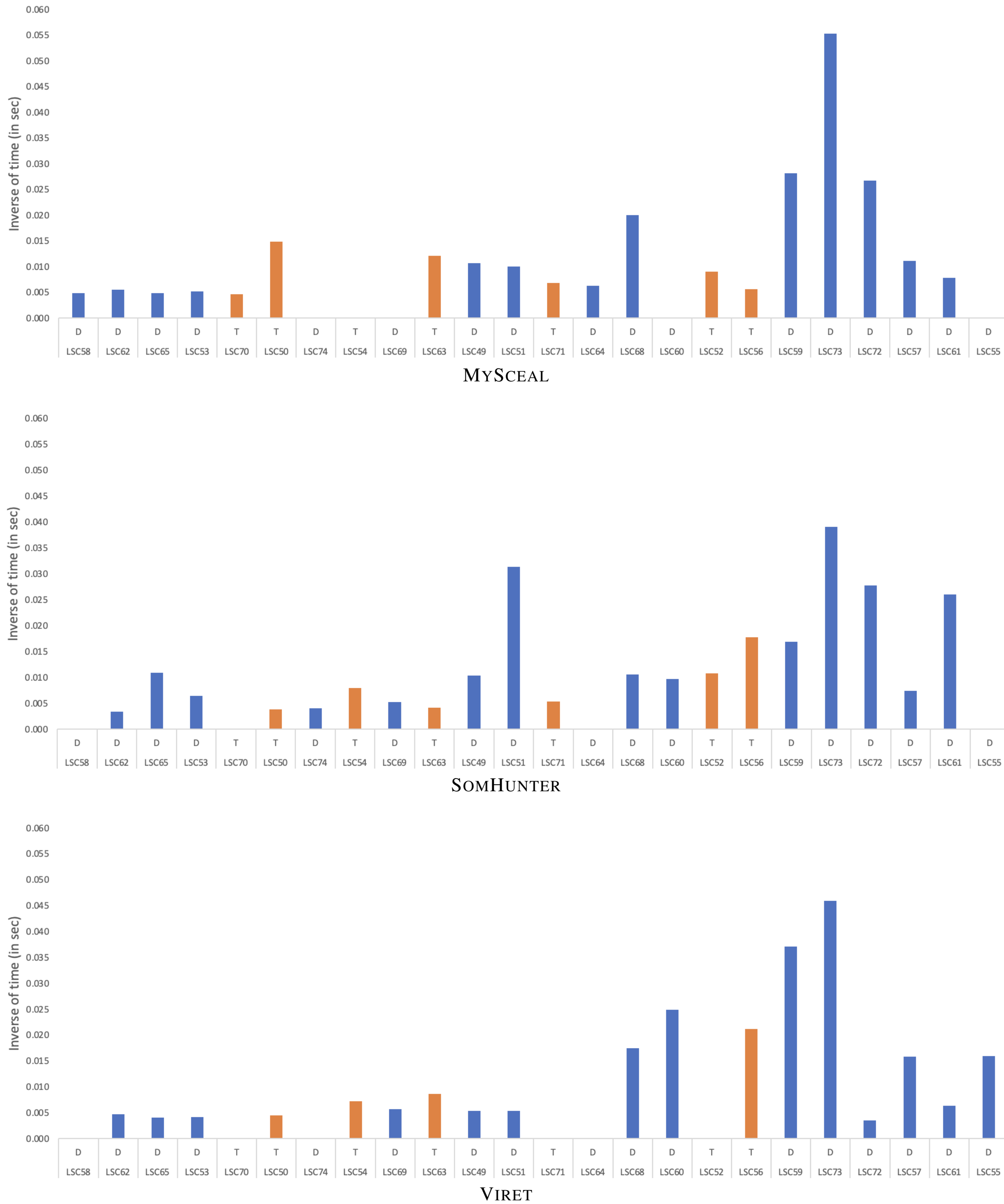


Figure 1: Inverse of time taken in seconds (sec) for MYSCEAL, SOMHUNTER and VIRET systems to retrieve the correct image across all topics; higher bar indicates speedier retrieval; empty bars indicate the correct image was not retrieved within the imposed time limit of 300 seconds; blue denotes a descriptive topic; orange denotes a temporal topic; topics are ordered from slowest overall topic (most difficult to retrieve) to quickest overall topic (least difficult to retrieve) based on average retrieval times of systems

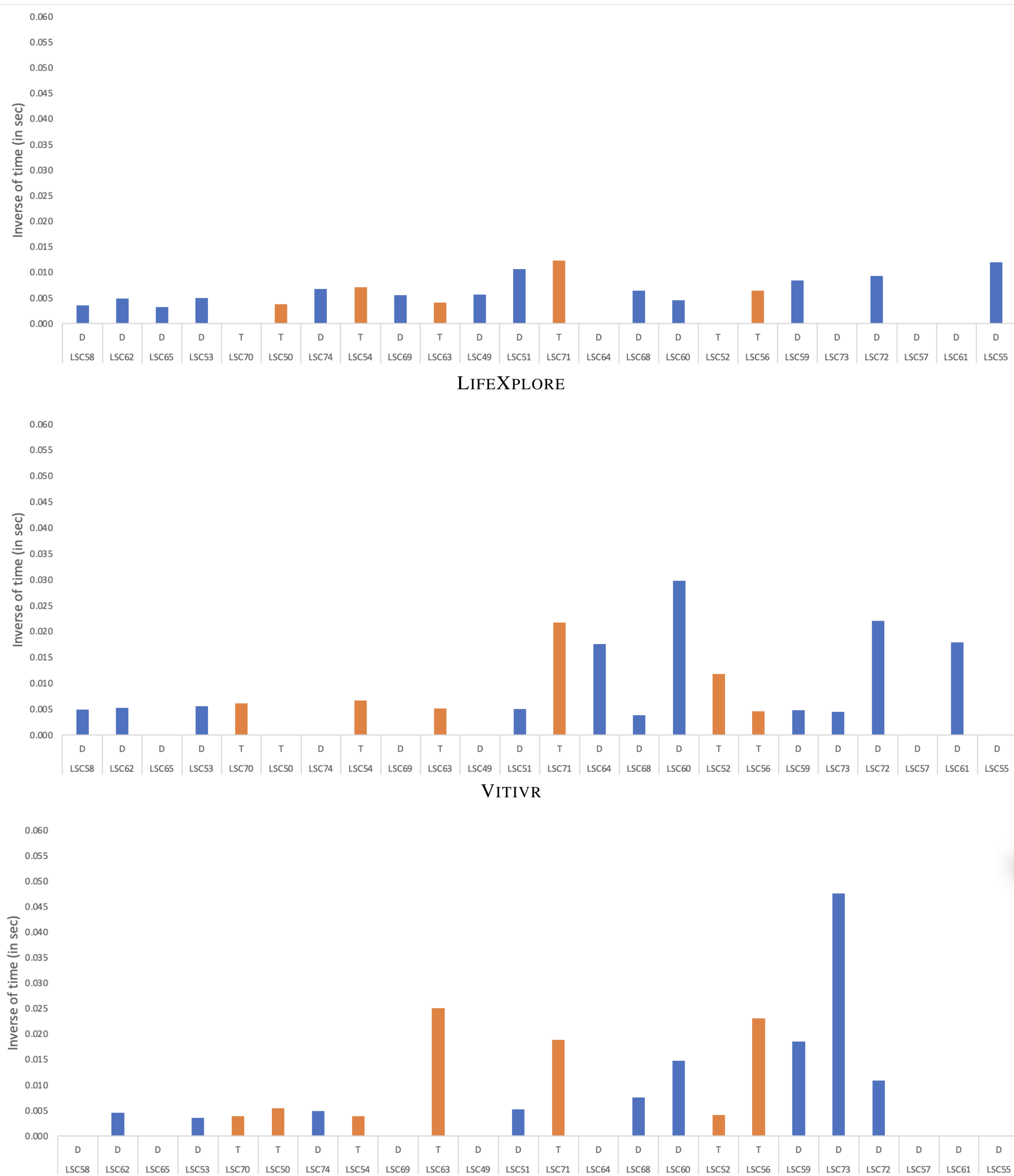
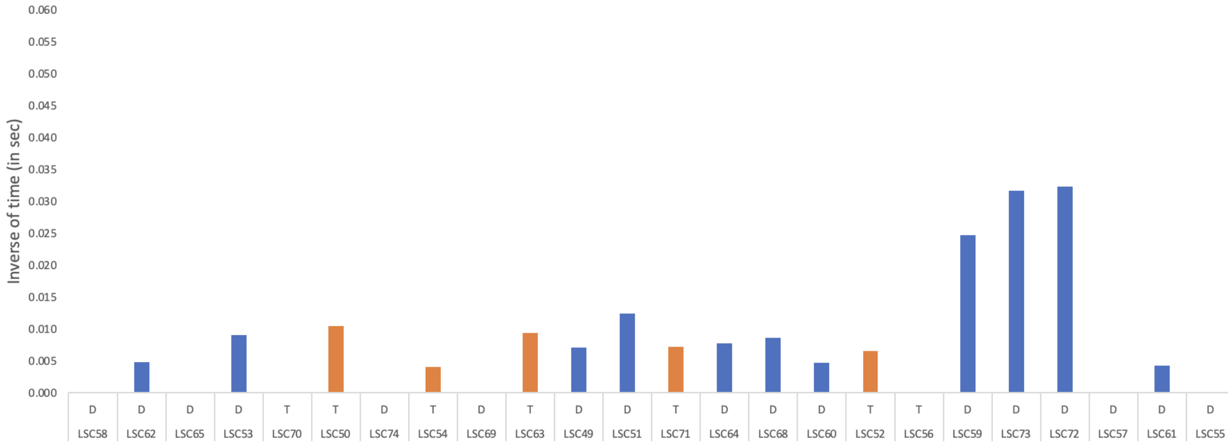
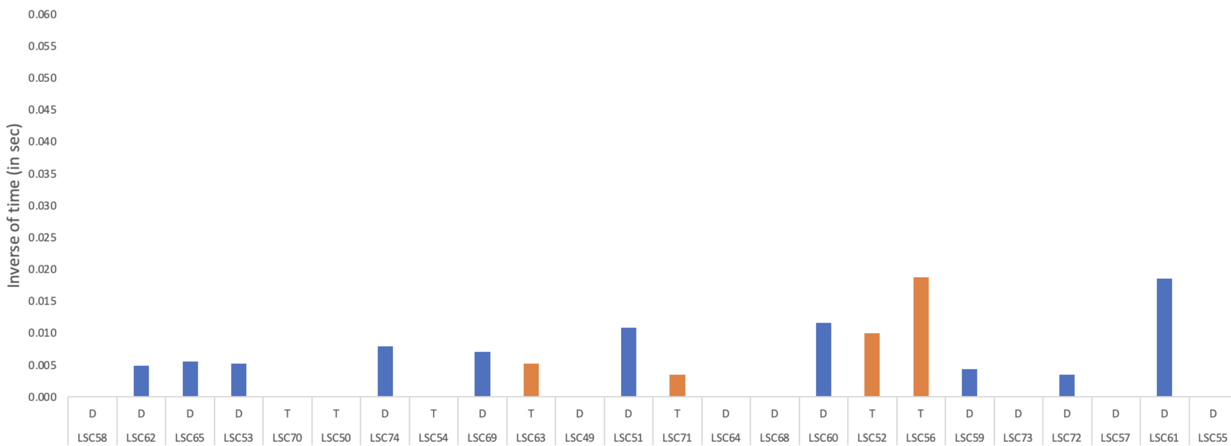


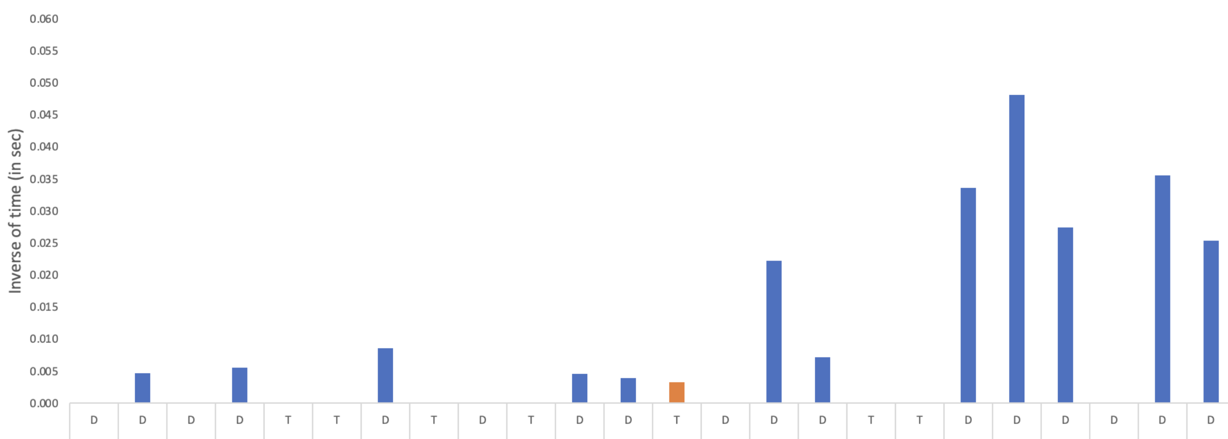
Figure 2: Inverse of time taken in seconds (sec) for LIFEXPLORE, VITIVR and EXQUISITOR systems to retrieve the correct image across all topics; higher bar indicates speedier retrieval; empty bars indicate the correct image was not retrieved within the imposed time limit of 300 seconds; blue denotes a descriptive topic; orange denotes a temporal topic; topics are ordered from slowest overall topic (most difficult) to quickest overall topic (least difficult to retrieve) based on average retrieval times of systems



VIRLE



HCMUS-FIRST



LIFESEEKER

Figure 3: Inverse of time taken in seconds (sec) for VIRLE, HCMUS-FIRST and LIFESEEKER systems to retrieve the correct image across all topics; higher bar indicates speedier retrieval; empty bars indicate the correct image was not retrieved within the imposed time limit of 300 seconds; blue denotes a descriptive topic; orange denotes a temporal topic; topics are ordered from slowest overall topic (most difficult) to quickest overall topic (least difficult to retrieve) based on average retrieval times of systems

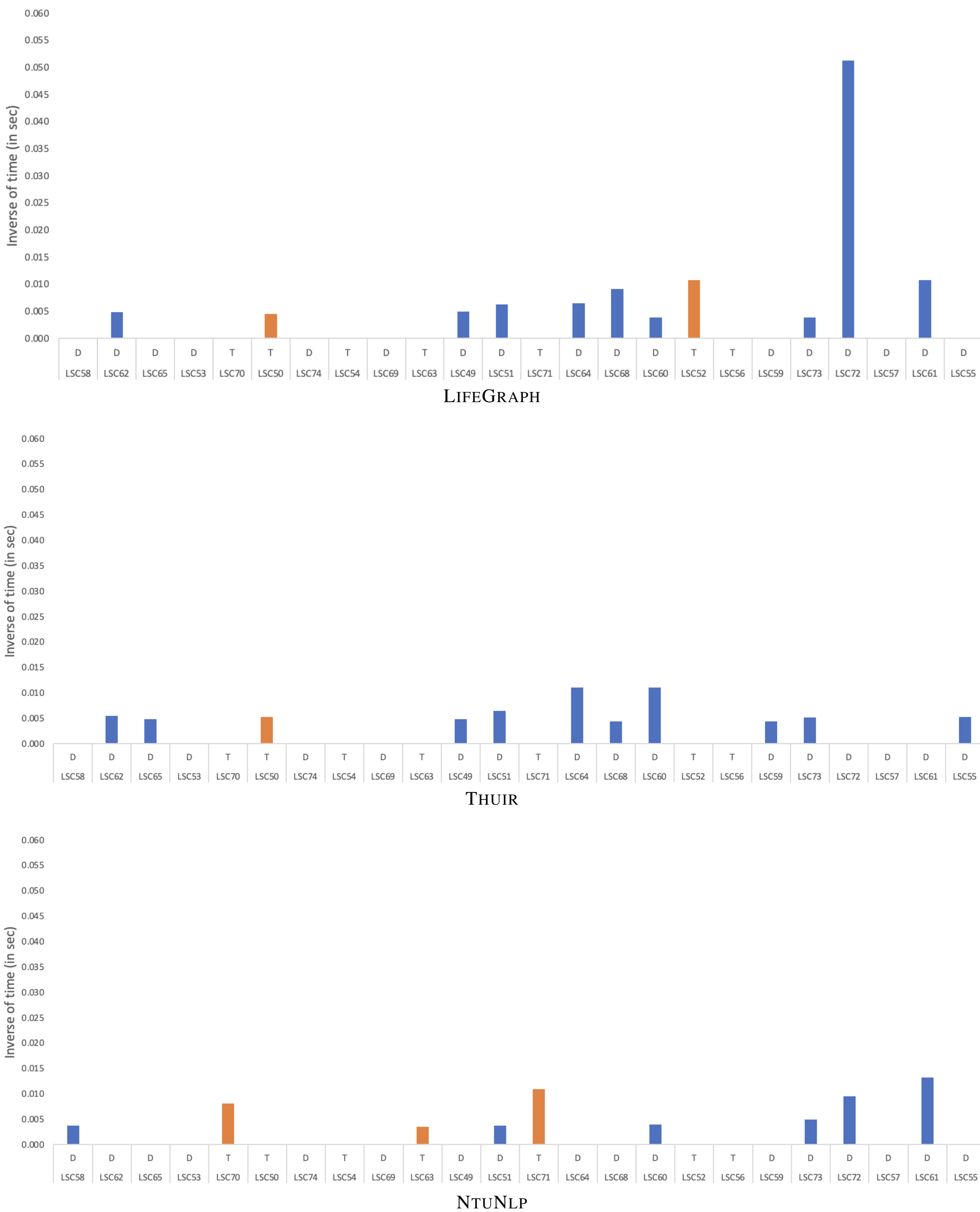


Figure 4: Inverse of time taken in seconds (sec) for LIFEGRAPH, THUIR and NTUNLP systems to retrieve the correct image across all topics; higher bar indicates speedier retrieval; empty bars indicate the correct image was not retrieved within the imposed time limit of 300 seconds; blue denotes a descriptive topic; orange denotes a temporal topic; topics are ordered from slowest overall topic (most difficult) to quickest overall topic (least difficult to retrieve) based on average retrieval times of systems

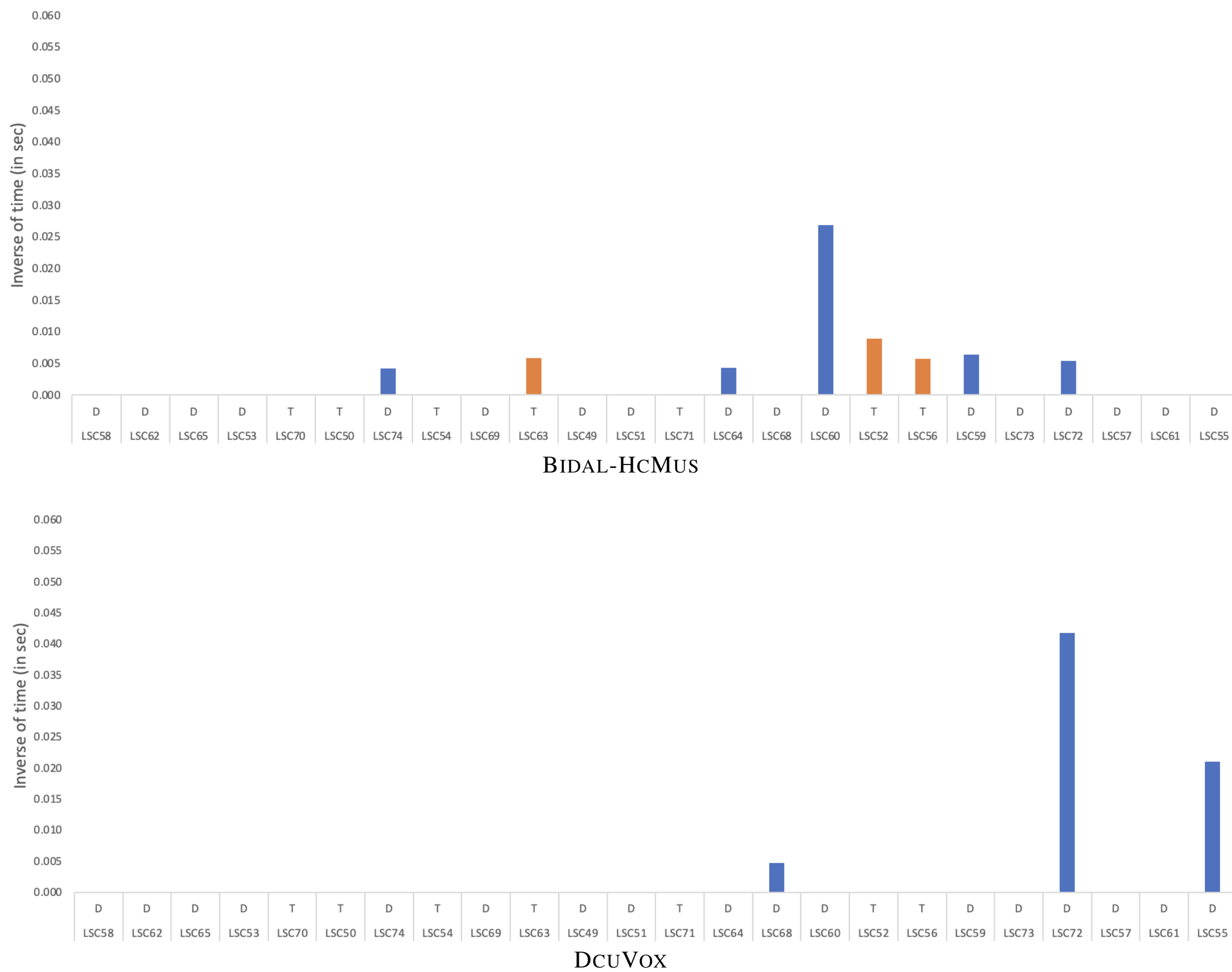


Figure 5: Inverse of time taken in seconds (sec) for BIDAL-HCMUS and DCUVOX systems to retrieve the correct image across all topics; higher bar indicates speedier retrieval; empty bars indicate the correct image was not retrieved within the imposed time limit of 300 seconds; blue denotes a descriptive topic; orange denotes a temporal topic; topics are ordered from slowest overall topic (most difficult) to quickest overall topic (least difficult to retrieve) based on average retrieval times of systems

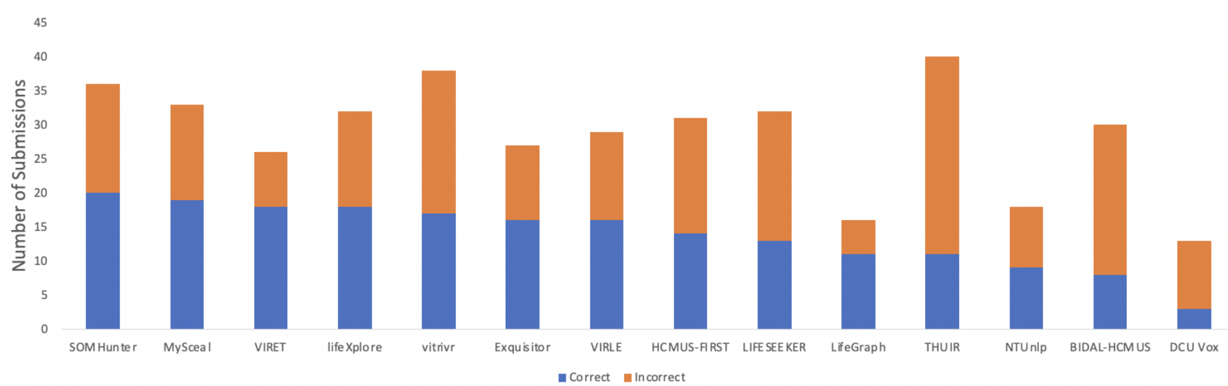


Figure 6: Number of Correct/Incorrect submissions across teams for all topics

1. Descriptive : Describes scene in greater detail (17 out of 24 topics)

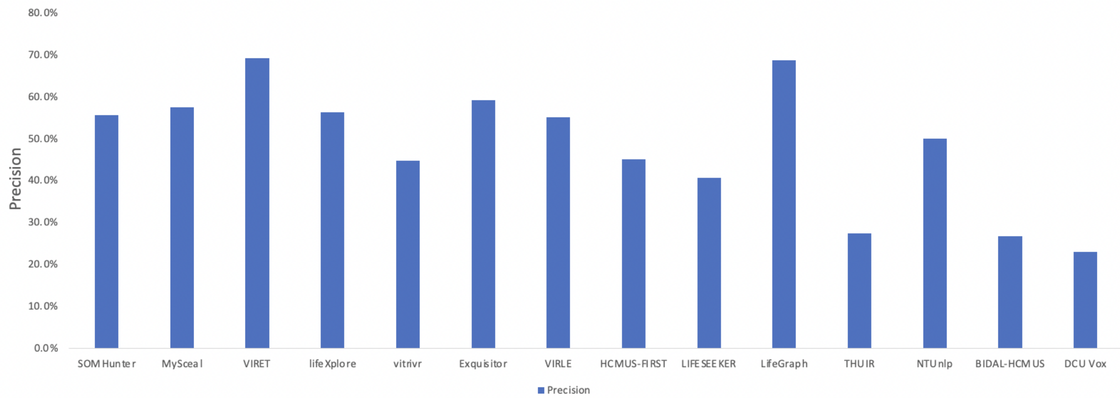


Figure 7: Precision (Correct submissions/Total submissions) across teams for all topics.

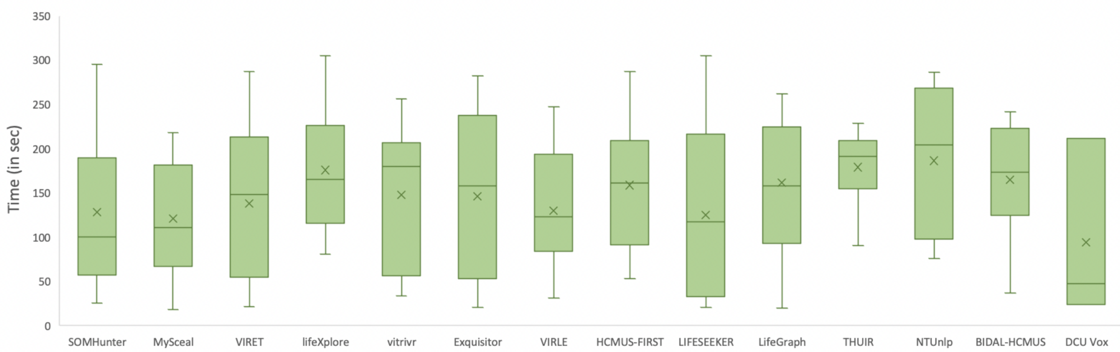


Figure 8: Time distribution (in sec) for correct submissions across teams for all topics.

2. Temporal : Puts more focus on temporal aspects than scene description (7 out of 24 topics).

The objective here is to analyze which systems are tuned to handle which type of topics.Figure 19 shows time distribution for correct response submission segregated on the basis of topic category. Figure 20 and 22 show Inverse of submission time for Temporal and Descriptive topics respectively for all teams.

We observe that systems like Exquisitor, MySceal, SOMHunter and vitrivr handle temporal queries extremely well while systems like DCU Vox, LifeSeeker, THUIR and NTUnlp perform poorly on these topics.

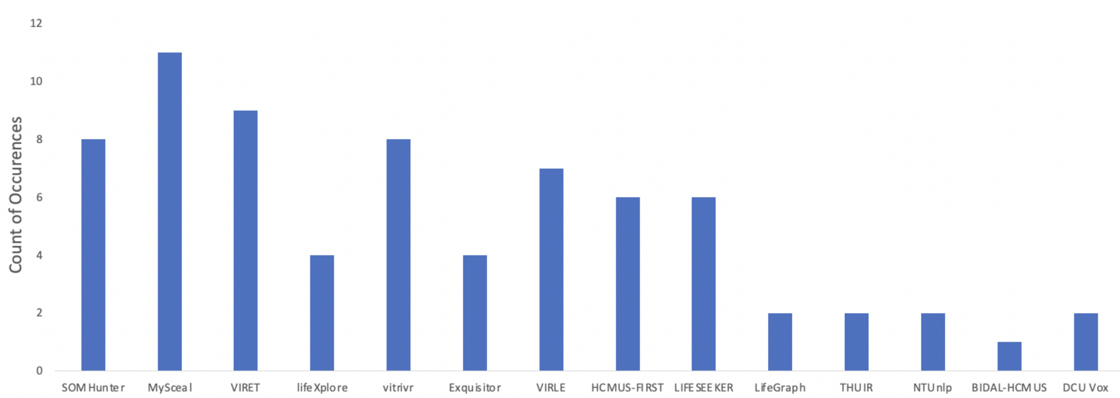


Figure 9: Count of occurrences when the system was in top-3 to submit a response across all topics.

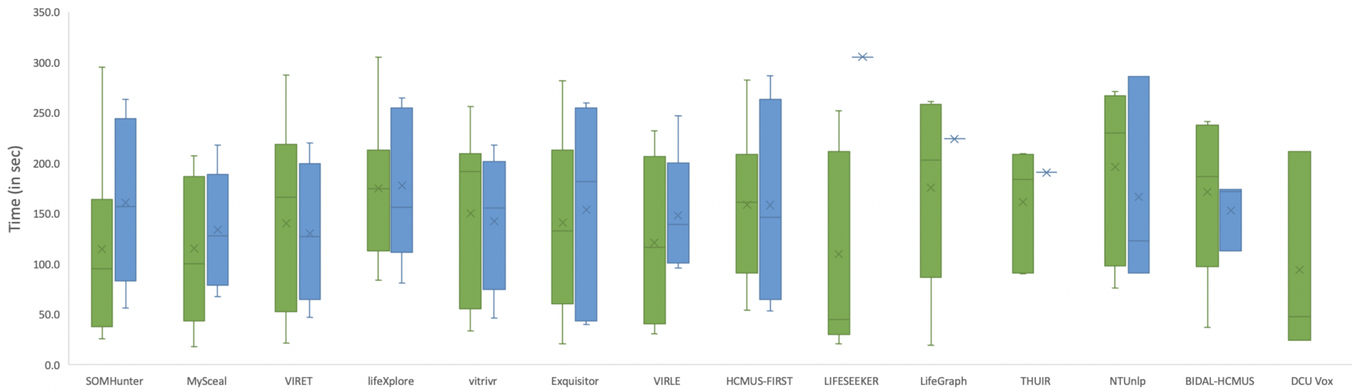


Figure 10: Time distribution (in sec) for correct submissions across teams for Descriptive and Temporal topics. Green bars represent descriptive while blue bars represent temporal temporal topics.

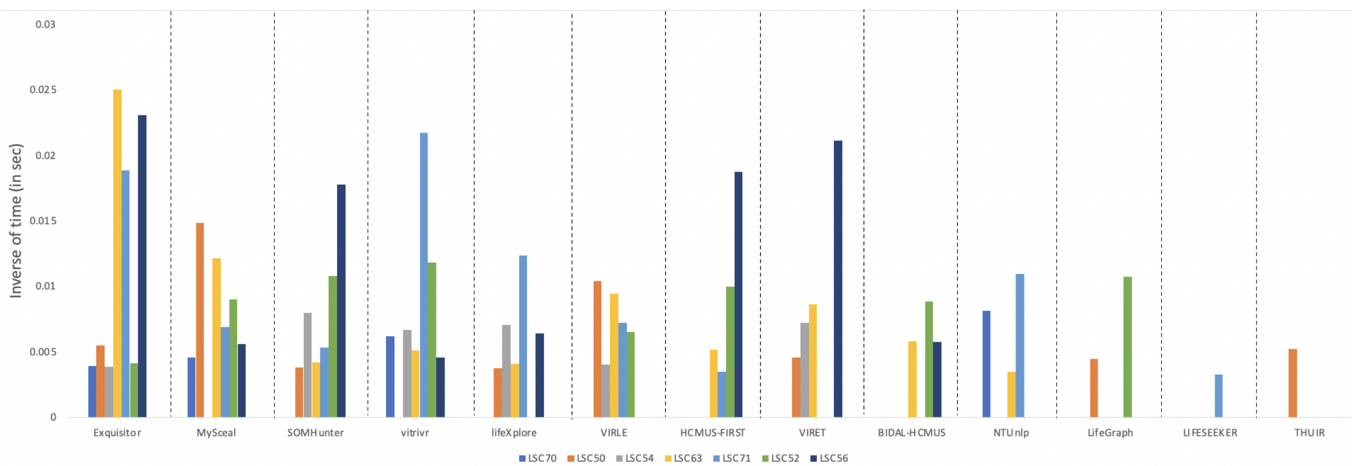


Figure 11: Inverse of time (in sec) for a correct response plotted against each team for temporal topics.Empty bars indicate unanswered topics. Higher bars indicate quicker correct submission.

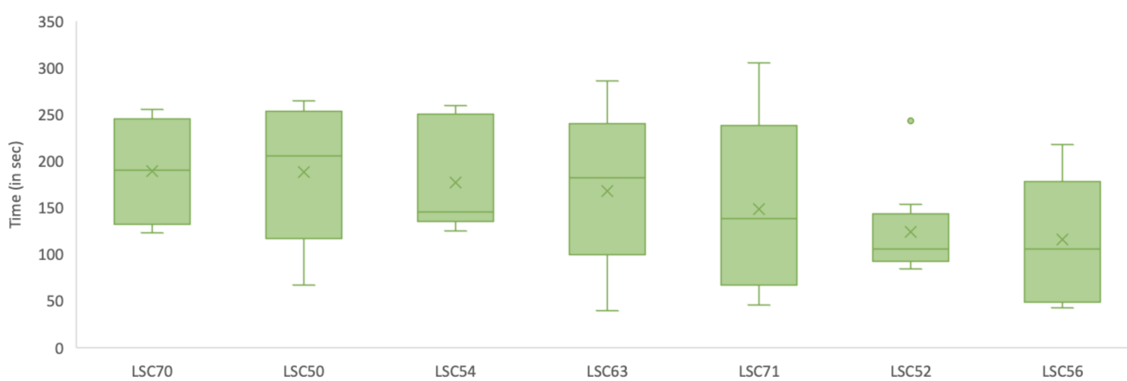


Figure 12: Time distribution (in sec) to submit a correct response for temporal topics.Topics are ordered left to right based on decreasing average submission time.

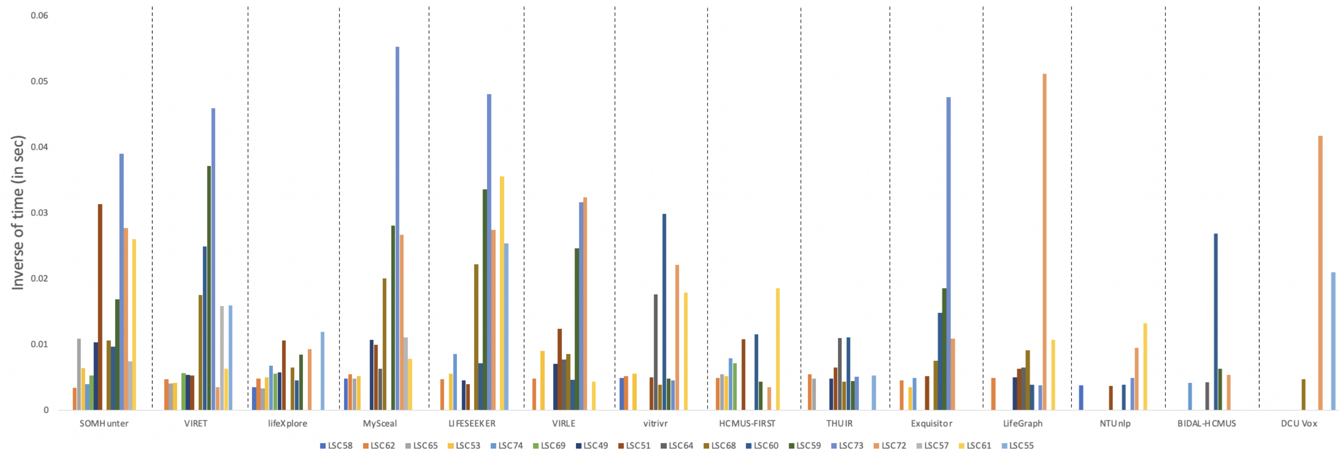


Figure 13: Inverse of time (in sec) for a correct response plotted for each team against descriptive topics.Empty bars indicate unanswered topics. Higher bars indicate quicker correct submission.

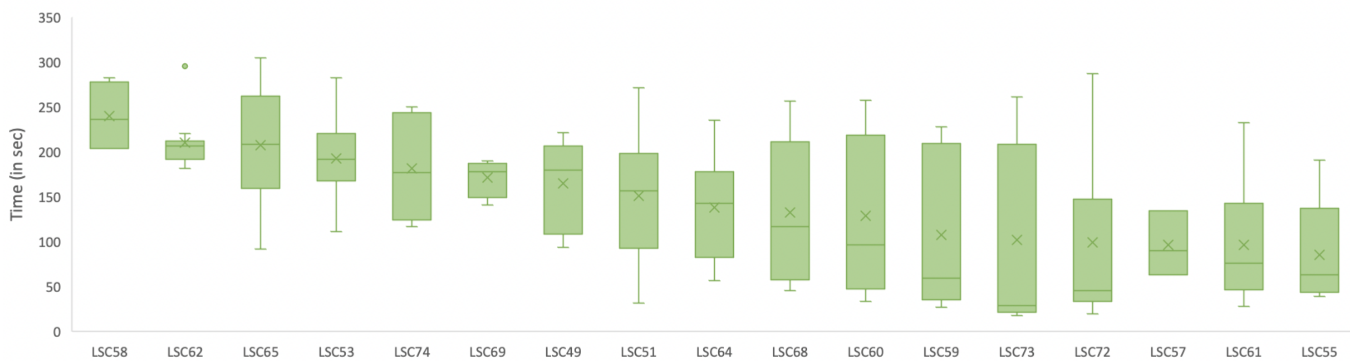


Figure 14: Time distribution (in sec) to submit a correct response for descriptive topics.Topics are ordered left to right based on decreasing average submission time.

Topics in Figure 20 and Figure 22 are ordered left to right on the basis of decreasing average submission time as shown in Figure 21 and Figure 23 respectively. A high average submission time may indicate more difficult topic while lower average time may indicate otherwise.

4 Conclusion

References

[1] Ly-Duyen Tran, Manh-Duy Nguyen, Nguyen Thanh Binh, Hyowon Lee, and Cathal Gurrin. Myscéal: An Experimental Interactive Lifelog Retrieval System for LSC'20. In *Proceedings of the Third Annual Workshop on Lifelog Search Challenge*, pages 23–28, Dublin Ireland, June 2020. ACM.

[2] František Mejzlík, Patrik Veselý, Miroslav Kratochvíl, Tomáš Souček, and Jakub Lokoč. SOMHunter for Lifelog Search. In *Proceedings of the Third Annual Workshop on Lifelog Search Challenge*, pages 73–75, Dublin Ireland, June 2020. ACM.

[3] Gregor Kovalčík, Vít Škrhak, Tomáš Souček, and Jakub Lokoč. VIRET Tool with Advanced Visual Browsing and Feedback. In *Proceedings of the Third Annual Workshop on Lifelog Search Challenge*, pages 63–66, Dublin Ireland, June 2020. ACM.

[4] Andreas Leibetseder and Klaus Schoeffmann. lifeXplore at the Lifelog Search Challenge 2020. In *Proceedings of the Third Annual Workshop on Lifelog Search Challenge*, pages 37–42, Dublin Ireland, June 2020. ACM.

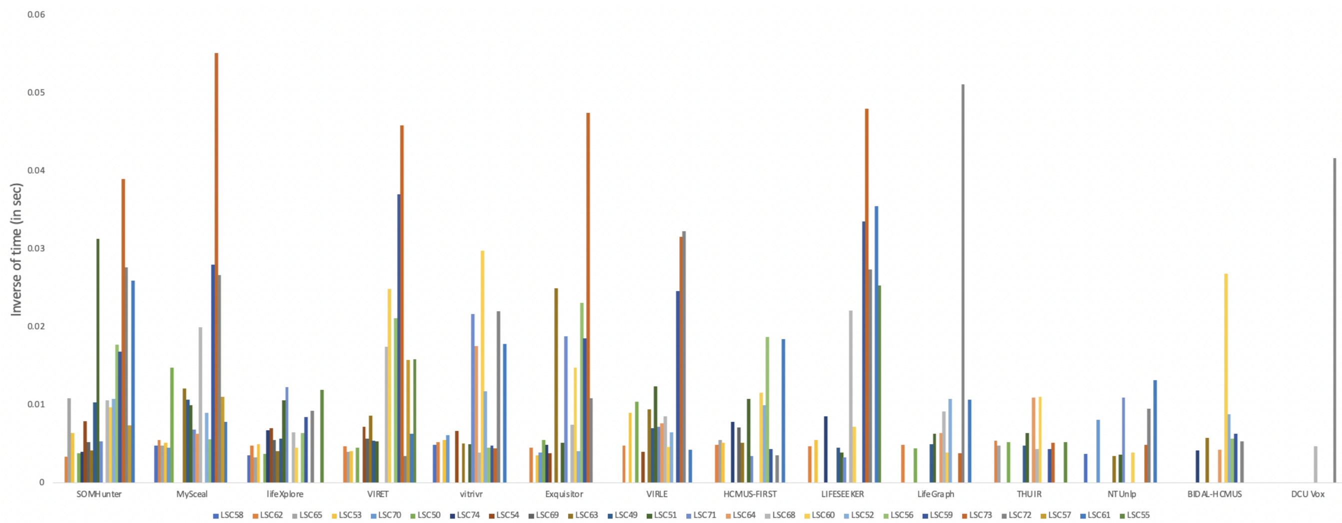


Figure 15: Inverse of time (in sec) for a correct response plotted for each team against all topics.Empty bars indicate unanswered topics. Higher bars indicate quicker correct submission.

[5] Silvan Heller, Mahnaz Amiri Parian, Ralph Gasser, Loris Sauter, and Heiko Schuldt. Interactive Lifelog Retrieval with vitrivr. In *Proceedings of the Third Annual Workshop on Lifelog Search Challenge*, pages 1–6, Dublin Ireland, June 2020. ACM.

[6] Omar Shahbaz Khan, Mathias Dybkjær Larsen, Liam Alex Sonto Poulsen, Björn Þór Jónsson, Jan Zahálka, Stevan Rudinac, Dennis Koelma, and Marcel Worring. Exquisitor at the Lifelog Search Challenge 2020. In *Proceedings of the Third Annual Workshop on Lifelog Search Challenge*, pages 19–22, Dublin Ireland, June 2020. ACM.

[7] Aaron Duane, Björn Þór Jónsson, and Cathal Gurrin. VRLE: Lifelog Interaction Prototype in Virtual Reality: Lifelog Search Challenge at ACM ICMR 2020. In *Proceedings of the Third Annual Workshop on Lifelog Search Challenge*, pages 7–12, Dublin Ireland, June 2020. ACM.

[8] Minh-Triet Tran, Thanh-An Nguyen, Quoc-Cuong Tran, Mai-Khiem Tran, Khanh Nguyen, Van-Tu Ninh, Tu-Khiem Le, Hoang-Phuc Trang-Trung, Hoang-Anh Le, Hai-Dang Nguyen, Trong-Le Do, Viet-Khoa Vo-Ho, and Cathal Gurrin. FIRST - Flexible Interactive Retrieval SysTem for Visual Lifelog Exploration at LSC 2020. In *Proceedings of the Third Annual Workshop on Lifelog Search Challenge*, pages 67–72, Dublin Ireland, June 2020. ACM.

[9] Tu-Khiem Le, Van-Tu Ninh, Minh-Triet Tran, Thanh-An Nguyen, Hai-Dang Nguyen, Liting Zhou, Graham Healy, and Cathal Gurrin. LifeSeeker 2.0: Interactive Lifelog Search Engine at LSC 2020. In *Proceedings of the Third Annual Workshop on Lifelog Search Challenge*, pages 57–62, Dublin Ireland, June 2020. ACM.

[10] Luca Rossetto, Matthias Baumgartner, Narges Ashena, Florian Ruosch, Romana Pernischová, and Abraham Bernstein. LifeGraph: A Knowledge Graph for Lifelogs. In *Proceedings of the Third Annual Workshop on Lifelog Search Challenge*, pages 13–17, Dublin Ireland, June 2020. ACM.

[11] Jiayu Li, Min Zhang, Weizhi Ma, Yiqun Liu, and Shaoping Ma. A Multi-level Interactive Lifelog Search Engine with User Feedback. In *Proceedings of the Third Annual Workshop on Lifelog Search Challenge*, pages 29–35, Dublin Ireland, June 2020. ACM.

[12] Tai-Te Chu, Chia-Chun Chang, An-Zi Yen, Hen-Hsen Huang, and Hsin-Hsi Chen. Multimodal Retrieval through Relations between Subjects and Objects in Lifelog Images. In *Proceedings of the Third Annual Workshop on Lifelog Search Challenge*, pages 51–55, Dublin Ireland, June 2020. ACM.

[13] Anh-Vu Mai-Nguyen, Trong-Dat Phan, Anh-Khoa Vo, Van-Luon Tran, Minh-Son Dao, and Koji Zettsu. BIDAL-HCMUS@LSC2020: An Interactive Multimodal Lifelog Retrieval with Query-to-Sample Attention-based Search Engine. In *Proceedings of the Third Annual Workshop on Lifelog Search Challenge*, pages 43–49, Dublin Ireland, June 2020. ACM.

[14] Ahmed Alateeq, Mark Roantree, and Cathal Gurrin. Voxento: A Prototype Voice-controlled Interactive Search Engine for Lifelogs. In *Proceedings of the Third Annual Workshop on Lifelog Search Challenge*, pages 77–81, Dublin Ireland, June 2020. ACM.

MySceal [1]	Proposed a modified version of term frequency-inverse document frequency (TF-IDF) to assess the relative importance of annotated objects within the image and a query mechanism to facilitate sequence queries to retrieve temporal events
SOMHunter [2]	An interactive retrieval tool designed for known-item and ad-hoc search tasks over image and video datasets. Their search strategy utilizes a text query initialization with optional temporal context and continues with refining the search with user relevance feedback on the intermediate result
VIRET [3]	A video retrieval system that focuses on result set presentation based on dynamically computed self-organized maps, as well as additional feedback about the distribution of relevance scores for a query.It supports multiple query mechanisms like query by text, query by drawing a color or semantic sketch and provides a dynamically constructed, semantically organized hierarchical browsing structure for top K relevant results.
LifeXplore [4]	A video exploration and retrieval tool which supports feature map browsing based on features like color, edge histogram and, visual concepts. It also supports multiple query mechanisms like query by concept/metadata, query by sketch, and query by similar images.
vitrivr [5]	An open source multimedia retrieval stack that supports the combination of queries across different modalities like query by example, query by sketch, textual queries or combinations. It also has support staged queries and temporal ordering in query formulation.
Exquisitor [6]	Uses relevance feedback from the user on media items to build a model (linear SVM) of the users' information needs. It combines efficient representation of data, an interactive classifier and also has support for temporal browsing of events.
VIRLE [7]	A virtual reality based information retrieval system for lifelogs and uses a gesture-based querying interface.The system includes features like event ranking and event visualization which is the grouping and ranking of semantically and temporally related real-life activities as 'events' and summarizing them into keyframes for efficient browsing.
HCMUS-FIRST [8]	Incorporates an integration platform to define and execute new query workflows, visualization layouts to visualize images into clusters with different semantic criteria, like color histograms, scene attributes, extracted deep features,etc. and has support for temporal navigation. It also uses image captions for better scene understanding and proposes an autoencoder-like method to measure the semantic relationship between query and image.
LifeSeeker [9]	It supports text query using a Bag-of-Words model with visual concept augmentation and also has support for temporal browsing of events using a time scaling factor.The system uses an interactive transitional graph-based filter to narrow events while transitioning from one place to another.
LifeGraph [10]	Supports retrieval by interlinking the lifelog data as a knowledge graph and also linking them to external, static knowledge bases in order to put the log as a whole as well as its individual entries into a broader context.It captures the internal relations of the various data modalities contained within a lifelog while simultaneously linking to large static knowledge bases like "Classification of Everyday Living" (COEL) and Wikidata in order to enrich the semantic context of the lifelog.
THUIR [11]	Incorporates multilevel features which are generated from lifelogs like image visual features, detail descriptions of the scene, and behavior expression to improve the search efficiency. The system supports multiple query mechanisms like query by text and query by image.
NTUnlp [12]	Encodes the information of relationships between subjects and objects in images by using a pre-trained relation graph generation model.It utilizes an end-to-end multi-level scene description network to construct the relation graphs between subjects and objects in images which is then projected to the embedding space for online retrieval.
BIDAL-HCMUS [13]	An interactive multimodal lifelog retrieval system with the query-to-sample attention-based search mechanism which utilizes deep learning techniques, clustering methods,and similarity search approaches to search for relevant images.
DCU Vox [14]	A voice-controlled interactive retrieval system for lifelogs. The system makes use of GoogleWeb Speech API to detect a spoken query and recognise specific commands such as submit a search task or stop recording. The search engine is implemented using MongoDB and Mongoose API.

Table 2: Summary of approaches taken by participants at LSC'20

	All Topics	Descriptive Topics	Temporal Topics
MySceal	79.1%	76.4%	85.7%
SOMhunter	83.3%	82.3%	85.7%
VIRET	75.0%	82.3%	57.1%
LifeXplore	75.0%	76.4%	71.4%
vitrivr	70.8%	64.7%	85.7%
Exquisitor	66.6%	52.9%	100.0%
VIRLE	66.6%	64.7%	71.4%
HCMUS-FIRST	58.3%	58.8%	57.1%
LifeSeeker	54.1%	70.5%	14.2%
LifeGraph	45.8%	52.9%	28.5%
THUIR	45.8%	58.8%	14.2%
NTUnlp	37.5%	35.2%	42.8%
BIDAL-HCMUS	33.3%	29.4%	42.8%
DCU Vox	12.0%	17.6%	0%

Table 3: Precision achieved by each system participating in LSC'20 for all topics (all), descriptive topics and temporal topics.