General Purpose Multimedia Retrieval with vitrivr at LSC'24

Loris Sauter
University of Basel
Basel, Switzerland
loris.sauter@unibas.ch

Ralph Gasser
University of Basel
Basel, Switzerland
ralph.gasser@unibas.ch

Laura Rettig
University of Basel
Basel, Switzerland
laura.rettig@unibas.ch

Heiko Schuldt
University of Basel
Basel, Switzerland
heiko.schuldt@unibas.ch

Luca Rossetto
University of Zurich
Zurich, Switzerland
rossetto@ifi.uzh.ch

ABSTRACT

The collection of lifelog data — visual and multi-sensory data, including biometric and spatiotemporal metadata — becomes easier and more supported by commercial products every year. Naturally, lifelog data is multi-modal, with arguably a major audio-visual component, such as captured videos, audio recordings and photos. For lifelog retrieval, the challenges of managing and accessing (visual) multimedia content are paired with the challenges of semi-structured and heterogeneous metadata. One approach to these challenges is the application of general-purpose, content-based multimedia retrieval in combination with traditional Boolean retrieval. In this paper, we present the latest iteration of vitrivr, a long-running participant in the Lifelog Search Challenge. After successfully replacing the retrieval engine Cineast with the vitrivr-engine for the structurally related Video Browser Showdown, we adjust the general purpose, content-based multimedia retrieval system to lifelog retrieval by extending the modular retrieval engine with Boolean retrieval and a model for metadata. In doing so, we continue to generalize the retrieval aspects also suitable for other applications and evaluate our system at the Lifelog Search Challenge 2024.

CCS CONCEPTS

• **Information systems** → **Search interfaces**; **Image search**; **Search interfaces**; **Image search**; **Users and interactive retrieval**; *Multimedia databases*; **Information retrieval**; • **Human-centered computing** → **Interactive systems and tools**.

KEYWORDS

Content-based Retrieval, Multimedia Retrieval, Lifelogging, Lifelog Search Challenge

ACM Reference Format:
Loris Sauter, Ralph Gasser, Laura Rettig, Heiko Schuldt, and Luca Rossetto. 2024. General Purpose Multimedia Retrieval with vitrivr at LSC'24. In *The 7th Annual ACM Lifelog Search Challenge (LSC '24), June 10, 2024, Phuket, Thailand*. ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3643489.3661120

LSC '24, June 10,2024, Phuket, Thailand

ACM ISBN 979-8-4007-0550-2/24/06
https://doi.org/10.1145/3643489.3661120

1 INTRODUCTION

Low-cost wearable devices and smartphones enable their wearer to generate an unprecedented amount of data, capturing various aspects of their life. The product of these activities is known as their individual *lifelog*, a multi-modal collection encompassing a rich spectrum of information, including but not limited to visual, geospatial, and biometric information. The size and, particularly, the heterogeneity of such a data collection, combined with the innate properties of queries on lifelog data, present a challenge to indexing, analysis, and retrieval of data items. The ubiquity of multi-modal data, such as in the form of lifelogs, motivates the development of robust information retrieval systems capable of handling data in diverse types and formats. A user trying to access an item in their personal lifelog may only partially remember facts about the event or context. As a consequence, it is necessary for a retrieval system to support different forms of querying that allow the user to combine their memory fragments. This ensures that the lifelog, created by automated data collection activities, can act as a form of augmented memory, enhancing an individual's ability to recall past experiences and events [18].

As part of the field of research around lifelog technologies, it is of great interest for the community to develop the technological capabilities necessary to exploit the potentials contained within lifelog collections. This research advances novel algorithms, systems, and user interfaces for multi-modal information retrieval and interactions with large data collections. Specifically, recent advancements have focused on improving the efficiency and intuitiveness of lifelog retrieval systems, with notable contributions in the areas of algorithmic development, multi-modal content descriptor models, system architecture, and user experience design [1, 2, 12, 13, 16, 23, 25, 26, 29, 30].

With this aim, the Lifelog Search Challenge (LSC) [9] serves as an annual evaluation of interactive lifelog retrieval systems held since 2018 [8], fostering collaboration and innovation in the field of lifelog retrieval systems. LSC provides a platform for researchers to showcase their advancements and compete on real-time search tasks. Motivated by and in execution similar to the Video Browser Showdown (VBS) [14], LSC is styled as a highly interactive competition, usually moderated, with live results (i.e., the competition score and ranking of the participants) being displayed using the evaluation system DRES [21]. Systems compete in real-time to solve search tasks in three categories [15] on an individual's lifelog as quickly as possible [7]:

Known-Item Search (KIS) Gradually more detailed textual hints describe one or more images from the dataset. There is a fixed pre-defined ground truth per such topic.

Ad-hoc Search (AS) Given a broad textual description, the goal is to find as many images as possible that fit this description. Human judges assess the submissions.

Question Answering (QA) Provided with a textual question, the aim is to find the answer in the lifelog data and submit it in textual form. Human judges assess the submissions.

Many systems have been proposed and adapted over the years to solve the lifelog search tasks. A common approach involves separate image and text encoders, yielding two sets of embeddings [2, 19]. However, recent advancements have seen an increase in the usage of combined image-text embedding spaces [12, 16, 30] through the emergence of CLIP and related tools to improve retrieval performance and speed. Memento 3.0 [1] improved the previous CLIP-powered system by clustering the dataset and reducing the search space and query matching to include only the best candidate cluster. MemoriEase [30] leverages the semantic search capabilities of ElasticSearch in conjunction with query expansion methods and the BLIP-2 embedding model to achieve high retrieval precision on a pre-processed subset of the collection. As a frequent participant in the challenge, the recent iteration of E-LifeSeeker [16] supports several embedding models, including CLIP and BLIP, while adding a Differential Networks approach to tackle QA tasks.

In this paper, we present the vitrivr[1] system with which we will participate in LSC'24. In contrast to last year's participation, we abandon the desktop-virtual reality hybrid system [26] and continue to use the newly introduced retrieval component, vitrivr-engine, as first proposed for VBS'24 [4]. In particular, we introduce more fine-grained query capabilities in terms of full-text search, (re-)introduce Boolean Retrieval into vitrivr-engine and, motivated by the structure of lifelog data, introduce a generalized metadata model, which can make use of vitrivr-engine's flexible data model.

2 RETRIEVAL WITH VITRIVR

vitrivr is an open-source[2], general-purpose content-based multimedia retrieval stack, supporting videos, images, temporally ordered image sequences and 3D models. Following a three-tier system approach as illustrated in Figure 2, vitrivr consists of the (i) a storage layer, (ii) a retrieval engine, and (iii) a frontend component. Over the past years, vitrivr has participated in LSC with different iterations of these components.

2.1 Participations in LSC

The inaugural participation's [20] components were **ADAM**$_{pro}$ [6] as storage engine, Cineast [22] as retrieval engine, and vitrivr-ng as frontend. Notably, that year, the vitrivr system won the LSC. Since the following year [10] — LSC'20 — the storage layer was replaced with Cottontail DB [5]. Starting from 2021, a virtual reality (VR) frontend — vitrivr VR [25] — complemented the desktop UI vitrivr-ng [11]. After two years of concurrent participation of the desktop [12] and VR-frontend [27], the two modalities were combined into an experimental, hybrid desktop-VR approach in last year's LSC.

[1]https://vitrivr.org/

[2]https://github.com/vitrivr/

One of the core strengths of the vitrivr system has been its modular architecture with regard to indexing and retrieval. Specifically, with minor adjustments (mostly in terms of configuration) the vitrivr system has been participating in VBS and LSC without major changes to the core system. This was possible due to the general-purpose retrieval methods available and the high level of configurability – both at query and indexing time. The retrieval methods include a plethora of query modes applicable to different retrieval scenarios, for instance both color and semantic sketch-based queries, text-based queries for OCR and deep learning-backed scene descriptions, Boolean queries on textual, numerical, and other values, geospatial queries, and queries that involve body pose. With the recent advancements in visual-text co-embedding-based methods, textual queries have become the dominant query modality, not only in the interactive video retrieval benchmark of VBS [14] but also in lifelog retrieval [28].

2.2 Introducing vitrivr-engine

The combined experiences of multiple consecutive years of interactive video retrieval evaluated at VBS, lifelog retrieval evaluated at LSC, and other applications of vitrivr, such as the projects XReco[3] and Archipanion[4], has led to the replacement of the retrieval component Cineast with its successor, vitrivr-engine for VBS'24. This step was necessitated by the fact that vitrivr is being used in various different contexts, which – despite having retrieval in common at their core – differ vastly in how data is being processed, represented, and searched. This is addressed by various changes, most importantly a more flexible *data model* and an (even more) freely configurable *indexing and querying pipeline*. In the following, we briefly summarise the model of vitrivr-engine, based on [4]:

vitrivr-engine's data model is centered around the concept of a *retrievable*, which represents any type of information that can be retrieved (e.g., an image file or a part thereof). These retrievable can be described by an arbitrary number of different types of *descriptors*, using an Entity-Attribute-Value (EAV) scheme. Such a descriptor can be anything, e.g., a scalar value, a vector, or some more complex structure. The different types of descriptors that can be used to describe a retrievable are defined as a *field* in the *schema* of a particular vitrivr-engine instance. Furthermore, arbitrary relationships between *retrievables* can be modeled using a graph structure.

Based on this data model, both *query execution* and *indexing* are then represented by operator pipelines that act on a stream of such *retrievables*. Essentially, a pipeline consists of a concatenation, dispersion, and aggregation of *operators*. While during indexing, operators are mainly used to decode and segment content from arbitrary sources (e.g., a series of video files) to then extract a range of feature descriptors, a querying pipeline uses operations to lookup values, expand relationships, calculate scores, or obtain aggregations of some sort. These high-level processing pipelines can be defined declaratively. In this model, querying and indexing are bridged by a type of operator called *analyser*, which is typically tied to a field and which extracts (from content during indexing) or retrieves (from the database during querying) feature descriptors.

[3]https://xreco.eu/

[4]https://dbis.dmi.unibas.ch/research/projects/archipanion/

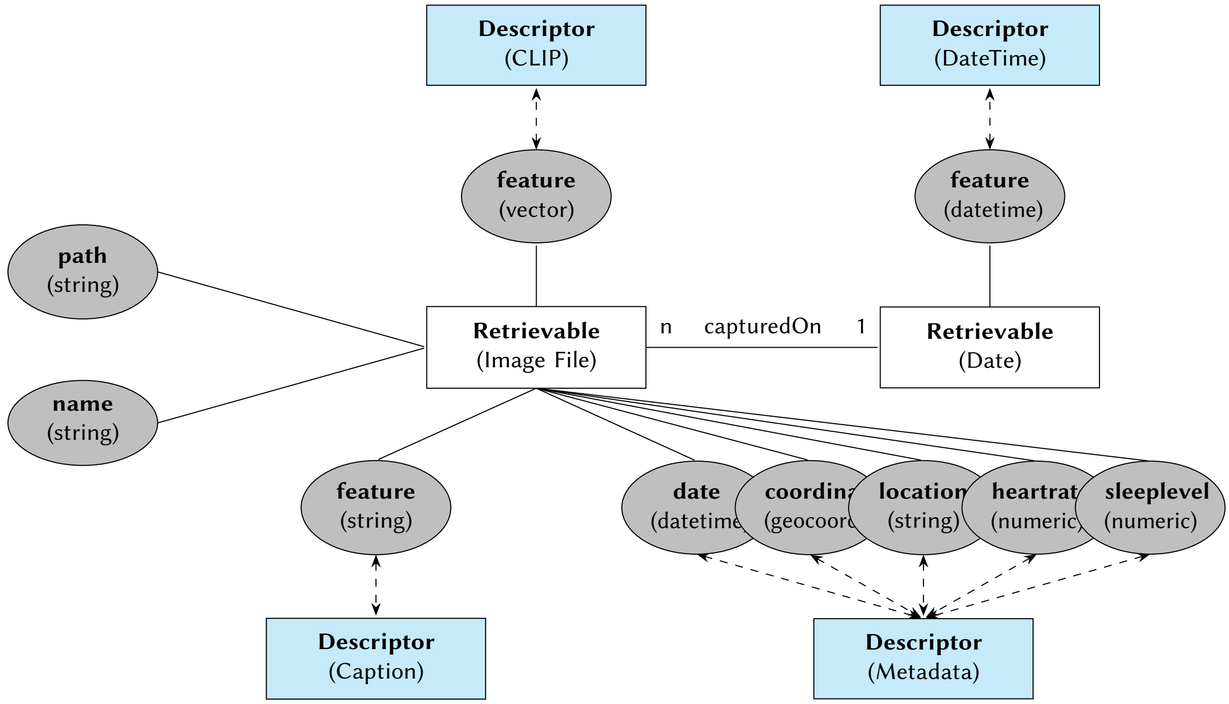

Figure 1: Simplified representation of the lifelog data in vitrivr-engine. The specific focus lies on the newly introduced metadata descriptor, which is a complex structure, consisting of multiple sub-fields. Another important aspect is the newly introduced datetime field type, with a corresponding descriptor. In order to have a readable graphic, we omitted the self-relations of the image file retrievable "before" and "after", both vital in representing the temporal order. The date retrievable groups the images of the same date together. While we extract the CLIP descriptor on our own, the caption is provided in the dataset. Other provided (textual) data are tags and OCR, which are omitted for readability reasons.

Consequently, vitrivr-engine can be adapted to a particular use case by configuring a schema to use types of fields that are relevant and useful. Both indexing and query pipelines can then be adapted based on this schema. This primarily means the selection of appropriate analysers relevant to lifelog retrieval. Of the general-purpose analysers vitrivr-engine currently provides, we consider the following to be particularly useful in the lifelog search setting: (i) Visual text co-embedding for scene description-based textual search provided by OpenCLIP [3], (ii) similarity search powered by DinoV2 [17], and (iii) generic full-text search in the provided textual data.

3 RETRIEVAL ON LIFELOG DATA

In addition to the previously introduced general-purpose retrieval methods, we introduce new concepts motivated by, but not exclusively for lifelog retrieval. Before, however, we briefly introduce the lifelog data used at LSC and then describe how this data is mapped to vitrivr-engine's data model. Subsequently, we address how retrieval is performed on this data and which modes of retrievals have been added to the most recent version of vitrivr-engine.

3.1 Lifelog Data

The LSC'24 dataset — collected over 18 consecutive months in 2019 and 2020 — consists of three separate but linked parts:

(1) The **Core Image Dataset** is the primary data to operate on, based on previous LSC analysis [24, 28], and consists of a set of 721 464 images that exhibit a 1024×768 pixels resolution, with redacted faces, text and certain activities. The medium for this part is the image, with the name of the image file corresponding to the date and time at which it was captured.
(2) The **Metadata** part provides a sparse table of available metadata such as spatial information with a semantic label, temporal information such as the time zone, sleep-related information such as the sleep level, as well as additional data such as music listened to and the heart rate at a specific time point. The medium for this part is text, of which some data can be transformed into semantically enriched types, such as date-time or coordinates.
(3) The **Visual Concepts** set contains OCR, generated tags, and captions, as well as the caption score on the non-redacted version of the imagery. The medium of this part is text.

The different parts of the data set can be connected using the time of capture. The individual images have the timestamp encoded in their filename. The metadata is captured every minute and can be associated with the images through their timestamp. The visual concepts are directly linked to the image they have been generated from by name. Since images may be captured more than once per

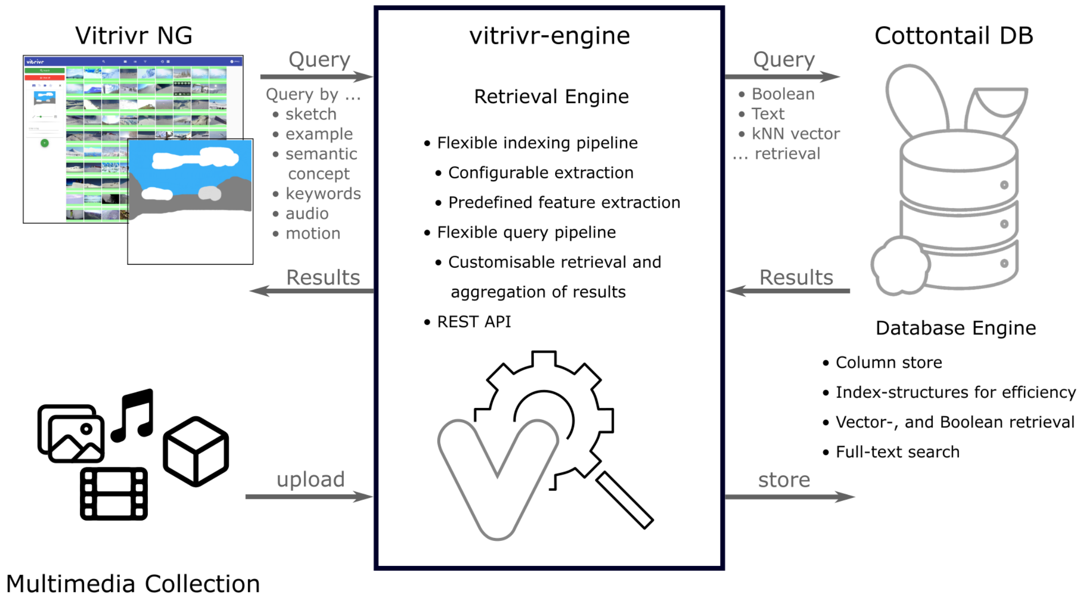

Figure 2: System architecture of vitrivr with vitrivr-engine and Cottontail DB. Both indexing and querying runtimes are specified as pipelines tailored to the needs of lifelog retrieval.

minute, hence metadata is more coarse-grained, there is a many-to-one relationship between the images to metadata. In addition to the relationships between those parts of the dataset, there is a strict temporal order on all data due to the nature of a lifelog. In Figure 3, we depict a sample image and the associated metadata.

3.2 Representing Lifelog Data in vitrivr-engine

At its core, the lifelog dataset consists of images, which can be retrieved using content-based methods. However, in lifelogs, these images also have a strict temporal order and exhibit temporal grouping, which in Cineast could only be reflected by introducing a specialized `IMAGE_SEQUENCE` media type. This was essentially a hard-coded change to the core Cineast system that embedded the dependency between the individual images and the date these images were taken in a rigid manner.

Due to the flexible data model of vitrivr-engine, there is no longer a need for such an artificial media type, as we can model the dependency between data and image as relationships between two retrievables, one representing a particular date and one representing the actual image. In doing so, we essentially follow the Knowledge-Graph-based approach of LifeGraph [23] at LSC'23. In contrast to LifeGraph, our graph relations are only used to model the temporal relationship between images. We argue that since lifelog data is rooted in temporal information, this approach is justifiable.

Motivated by the very successful participation of lifeXplore [24] in 2023 and their findings, we describe every lifelog image by extracting feature vectors to embed the image's content in a visual-text co-embedding space using OpenCLIP [3], which is a feature descriptor bundled with vitrivr-engine. In addition, we also incorporate the provided metadata and visual concepts. In the terminology of vitrivr-engine, we configure *fields* for all of this data on the *LSC* schema. That is, there are OpenCLIP and DinoV2 vector-based fields and text-based fields for the provided visual concepts. In addition, we introduce a new *struct descriptor*, specifically designed for the

Figure 3: Example lifelog image 20200106_035129 from the perspective of the lifelogger with redacted faces. The provided caption for this photo is "`a group of people sitting at a table with food and drinks`". Metadata for this specific moment are sparse, however temporally close (30 minutes), we learn that this was at the semantic location of "`South Korea, South Korea, South Korea`", as the metadata entry reveals.

provided metadata, which essentially is a list of key-value pairs. We use a field that is backed by this new descriptor for the provided metadata, excluding provided OCR, scene captions and tags. To accommodate the temporal data, we also introduce a new descriptor to handle dates, for which we configure a field as well. We have illustrated a simplified version of the data model in Figure 1. In order to de-clutter the graphic, we only render a subset of the fields with a focus on newly introduced representation methods.

3.3 Retrieving Lifelog Data in vitrivr-engine

Based on the experience of previous LSC installments [24, 28], co-embeddings are most beneficial for visual lifelog retrieval. Such queries are received as text in vitrivr-engine and then transformed into a vector in order to perform a nearest neighbor search, a feature already present in vitrivr-engine. Equally present in vitrivr-engine is full-text search on textual fields, e.g., the provided captions.

However, what has been missing so far and is going to be added for this iteration is the ability to combine such content-based search methods with *Boolean predicates* that constrain the results to entries that match certain boundary conditions. An example for such a query would be only to consider items that were taken on a Monday (in addition to having a textual description of a scene). Such conditions can be translated to very simple Boolean queries on scalar descriptors and/or parts of more complex descriptors - so-called *sub-fields*.

With the addition of Boolean predicates on the level of sub-fields, we also relax the existing query model such that any query might target a sub-field. Experience has shown that this can be beneficial for other query types as well, for instance full-text search, also in other contexts than lifelog retrieval.

4 IMPLEMENTATION

The vitrivr implementation and deployment at LSC'24 follows the traditional three-tier architecture found in earlier installments [12, 20, 26] and will be similar to the one used at VBS'24 [4]. The architecture is depicted in Figure 2. The two most notable differences in contrast to last year's participation [26] will be the discontinuation of the hybrid system and the employment of the new retrieval engine, vitrivr-engine. As a storage layer, we still use the latest version of Cottontail DB [5].

4.1 vitrivr-engine

The new vitrivr-engine is the successor of the feature extraction and retrieval engine Cineast and was first introduced in [4]. It has been written in Kotlin and runs in the Java Virtual Machine (JVM), that is, it can be used independently of the platform.

vitrivr-engine follows a few fundamental design decisions necessitated by it being used in various different contexts, which all do have retrieval in common yet differ in how data is being processed, represented, and searched. These different requirements are supported by introducing the following consequences:

Most importantly, the new vitrivr-engine decouples the high-level concepts from the concrete implementations much stricter than its predecessor Cineast. Different aspects, such as database connectivity, analysers and other types of operators that nor not strictly necessary to operate the core components, are bundled into different modules that can or cannot be used. Consequently, vitrivr-engine can be used in projects either in a very minimalist configuration where the user provides all the logic or using a standard, general-purpose setup.

This is made possible by a plugin architecture that allows a user of vitrivr-engine to provide custom implementation for virtually every component. This includes (but is not limited) to all types of operators used during querying and indexing, logic to map these operations to a persistence layer (which can be Cottontail DB [5] but could also be something else), and all logic to resolve and access external resources.

4.2 vitrivr-engine at LSC

The analysers used during LSC– namely OpenCLIP and DinoV2– are part of vitrivr-engine's standard features module. The extraction is facilitated by an external as a standalone Python server, with which it communicates via a RESTful API. The additions that facilitate Boolean search will be added to vitrivr-engine's core module.

To describe query pipelines, vitrivr-engine also provides a RESTful API, i.e., the wire-format for these queries is JSON, as seen in Listing 1. The frontend will facilitate query formulation by the user and then translate the to this low-level query language.

Listing 1: An example query for two inputs, a textual input to be processed by OpenCLIP and a Boolean query on a sub-field.

```
{
"inputs":{
    "i_clip": {"type": "TEXT", "data": "group of people with food"},
    "i_month": {"type": "TEXT", "data": "JANUARY",
        "comparison": "=="}
  },
  "operations":{
    "clip": {"type": "RETRIEVER", "field": "clip",
        "input": "i_clip"},
    "boolean": "{"type":"RETRIEVER", "field": "date.month",
        "input": "i_month"}
  }
}
```

5 CONCLUSION

In this paper, we have presented the general purpose, open-source, multi-modal multimedia retrieval stack vitrivr, particularly its new retrieval engine vitrivr-engine, with which we will participate at LSC'24. Specifically, we presented how we adjusted the general-purpose retrieval data model to accommodate lifelog data, particularly metadata and temporal data. In order to tackle the lifelog challenge, we added Boolean retrieval and full-text search on a more fine-grained level than previously existing in vitrivr-engine. Apart from the provided visual concepts and motivated by last year's winner, lifeXplore, we employ OpenCLIP as our main image retrieval method.

ACKNOWLEDGMENTS

This work was partly supported by the Swiss National Science Foundation through project "MediaGraph" (contract no. 202125), the Swiss State Secretariat for Education, Research and Innovation through the Horizon Europe project "XReco" (contract no. 22.00268) and Innosuisse through "Archipanion" (contract no. 104.460 IP-ICT).

REFERENCES

[1] Naushad Alam, Yvette Graham, and Cathal Gurrin. 2023. Memento 3.0: An Enhanced Lifelog Search Engine for LSC'23. In *Proceedings of the 6th Annual ACM Lifelog Search Challenge, LSC 2023, Thessaloniki, Greece, June 12-15, 2023*. ACM, 41–46. https://doi.org/10.1145/3592573.3593103

[2] Ahmed Alateeq, Mark Roantree, and Cathal Gurrin. 2023. Voxento 4.0: A More Flexible Visualisation and Control for Lifelogs. In *Proceedings of the 6th Annual ACM Lifelog Search Challenge* (Thessaloniki, Greece) *(LSC '23)*. Association for Computing Machinery, New York, NY, USA, 7–12. https://doi.org/10.1145/3592573.3593097

[3] Mehdi Cherti, Romain Beaumont, Ross Wightman, Mitchell Wortsman, Gabriel Ilharco, Cade Gordon, Christoph Schuhmann, Ludwig Schmidt, and Jenia Jitsev. 2023. Reproducible Scaling Laws for Contrastive Language-Image Learning. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2023, Vancouver, BC, Canada, June 17-24, 2023*. IEEE, 2818–2829. https://doi.org/10.1109/CVPR52729.2023.00276

[4] Ralph Gasser, Rahel Arnold, Fynn Faber, Heiko Schuldt, Raphael Waltenspül, and Luca Rossetto. 2024. A New Retrieval Engine for Vitrivr. In *MultiMedia Modeling*, Stevan Rudinac, Alan Hanjalic, Cynthia Liem, Marcel Worring, Björn Þór Jónsson, Bei Liu, and Yoko Yamakata (Eds.). Vol. 14557. Springer Nature Switzerland, Cham, 324–331. https://doi.org/10.1007/978-3-031-53302-0_28

[5] Ralph Gasser, Luca Rossetto, Silvan Heller, and Heiko Schuldt. 2020. Cottontail DB: An Open Source Database System for Multimedia Retrieval and Analysis. In *Proceedings of the 28th ACM International Conference on Multimedia (MM '20)*. Association for Computing Machinery, New York, NY, USA, 4465–4468. https://doi.org/10.1145/3394171.3414538

[6] Ivan Giangreco and Heiko Schuldt. 2016. ADAM pro : Database Support for Big Multimedia Retrieval. *Datenbank-Spektrum* 16, 1 (March 2016), 17–26. https://doi.org/10.1007/s13222-015-0209-y

[7] 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 *Proceedings of the 2021 International Conference on Multimedia Retrieval*. ACM, Taipei Taiwan, 690–691. https://doi.org/10.1145/3460426.3470945

[8] Cathal Gurrin, Klaus Schoeffmann, Hideo Joho, Duc-Tien Dang-Nguyen, Michael Riegler, and Luca Piras (Eds.). 2018. *LSC'18: Proceedings of the 2018 ACM Workshop on the Lifelog Search Challenge : June 11, 2018, Yokohama, Japan*. The Association for Computing Machinery, New York, New York.

[9] Cathal Gurrin, Liting Zhou, Graham Healy, Werner Bailer, Duc-Tien Dang-Nguyen, Steve Hodges, Björn Þór, Jakub Lokoc, Luca Rossetto, Minh-Triet Tran, and Klaus Schoeffmann. 2024. Introduction to the Seventh Annual Lifelog Search Challenge, LSC'24. In *Proceedings of the 2024 International Conference on Multimedia Retrieval (ICMR'24)* (Phuket, Thailand) *(ICMR '24)*. Association for Computing Machinery, New York, NY, USA, 2. https://doi.org/10.1145/3652583.3658891

[10] Silvan Heller, Mahnaz Amiri Parian, Ralph Gasser, Loris Sauter, and Heiko Schuldt. 2020. Interactive Lifelog Retrieval with Vitrivr. In *Proceedings of the Third Annual Workshop on Lifelog Search Challenge*. ACM, Dublin Ireland, 1–6. https://doi.org/10.1145/3379172.3391715

[11] Silvan Heller, Ralph Gasser, Mahnaz Parian-Scherb, Sanja Popovic, Luca Rossetto, Loris Sauter, Florian Spiess, and Heiko Schuldt. 2021. Interactive Multimodal Lifelog Retrieval with Vitrivr at LSC 2021. In *Proceedings of the 4th Annual on Lifelog Search Challenge*. ACM, Taipei Taiwan, 35–39. https://doi.org/10.1145/3463948.3469062

[12] Silvan Heller, Luca Rossetto, Loris Sauter, and Heiko Schuldt. 2022. Vitrivr at the Lifelog Search Challenge 2022. In *Proceedings of the 5th Annual on Lifelog Search Challenge*. ACM, Newark NJ USA, 27–31. https://doi.org/10.1145/3512729.3533003

[13] Maria Tysse Hordvik, Julie Sophie Teilstad Østby, Manoj Kesavulu, Thao-Nhu Nguyen, Tu-Khiem Le, and Duc-Tien Dang-Nguyen. 2023. LifeLens: Transforming Lifelog Search with Innovative UX/UI Design. In *Proceedings of the 6th Annual ACM Lifelog Search Challenge, LSC 2023, Thessaloniki, Greece, June 12-15, 2023*. ACM, 1–6. https://doi.org/10.1145/3592573.3593096

[14] Jakub Lokoč, Stelios Andreadis, Werner Bailer, Aaron Duane, Cathal Gurrin, Zhixin Ma, Nicola Messina, Thao-Nhu Nguyen, Ladislav Peška, Luca Rossetto, Loris Sauter, Konstantin Schall, Klaus Schoeffmann, Omar Shahbaz Khan, Florian Spiess, Lucia Vadicamo, and Stefanos Vrochidis. 2023. Interactive Video Retrieval in the Age of Effective Joint Embedding Deep Models: Lessons from the 11th VBS. *Multimedia Systems* (Aug. 2023), 3481–3504. https://doi.org/10.1007/s00530-023-01143-5

[15] Jakub Lokoč, Werner Bailer, Kai Uwe Barthel, Cathal Gurrin, Silvan Heller, Björn Þór Jónsson, Ladislav Peška, Luca Rossetto, Klaus Schoeffmann, Lucia Vadicamo, Stefanos Vrochidis, and Jiaxin Wu. 2022. A Task Category Space for User-Centric Comparative Multimedia Search Evaluations. In *MultiMedia Modeling*, Björn Þór Jónsson, Cathal Gurrin, Minh-Triet Tran, Duc-Tien Dang-Nguyen, Anita Min-Chun Hu, Binh Huynh Thi Thanh, and Benoit Huet (Eds.). Vol. 13141. Springer International Publishing, Cham, 193–204. https://doi.org/10.1007/978-3-030-98358-1_16

[16] Thao-Nhu Nguyen, Tu-Khiem Le, Van-Tu Ninh, Cathal Gurrin, Minh-Triet Tran, Nguyen Thanh Binh, Graham Healy, Annalina Caputo, and Sinéad Smyth. 2023. E-LifeSeeker: An Interactive Lifelog Search Engine for LSC'23. In *Proceedings of the 6th Annual ACM Lifelog Search Challenge, LSC 2023, Thessaloniki, Greece, June 12-15, 2023*. ACM, 13–17. https://doi.org/10.1145/3592573.3593098

[17] Maxime Oquab, Timothée Darcet, Théo Moutakanni, Huy Vo, Marc Szafraniec, Vasil Khalidov, Pierre Fernandez, Daniel Haziza, Francisco Massa, Alaaeldin El-Nouby, Mahmoud Assran, Nicolas Ballas, Wojciech Galuba, Russell Howes, Po-Yao Huang, Shang-Wen Li, Ishan Misra, Michael G. Rabbat, Vasu Sharma, Gabriel Synnaeve, Hu Xu, Hervé Jégou, Julien Mairal, Patrick Labatut, Armand Joulin, and Piotr Bojanowski. 2023. DINOv2: Learning Robust Visual Features without Supervision. *CoRR* abs/2304.07193 (2023). https://doi.org/10.48550/ARXIV.2304.07193 arXiv:2304.07193

[18] Ricardo Ribeiro, Alina Trifan, and António JR Neves. 2022. Lifelog retrieval from daily digital data: narrative review. *JMIR mHealth and uHealth* 10, 5 (2022), e30517.

[19] Ricardo F. Ribeiro, Luísa Amaral, Wei Ye, Alina Trifan, António J. R. Neves, and Pedro Iglésias. 2023. MEMORIA: A Memory Enhancement and MOment RetrIeval Application for LSC 2023. In *Proceedings of the 6th Annual ACM Lifelog Search Challenge, LSC 2023, Thessaloniki, Greece, June 12-15, 2023*. ACM, 18–23. https://doi.org/10.1145/3592573.3593099

[20] 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*. ACM, Ottawa ON Canada, 27–31. https://doi.org/10.1145/3326460.3329160

[21] Luca Rossetto, Ralph Gasser, Loris Sauter, Abraham Bernstein, and Heiko Schuldt. 2021. A System for Interactive Multimedia Retrieval Evaluations. In *MultiMedia Modeling*, Jakub Lokoč, Tomáš Skopal, Klaus Schoeffmann, Vasileios Mezaris, Xirong Li, Stefanos Vrochidis, and Ioannis Patras (Eds.), Vol. 12573. Springer International Publishing, Cham, 385–390. https://doi.org/10.1007/978-3-030-67835-7_33

[22] Luca Rossetto, Ivan Giangreco, Claudiu Tanase, and Heiko Schuldt. 2016. Vitrivr: A Flexible Retrieval Stack Supporting Multiple Query Modes for Searching in Multimedia Collections. In *Proceedings of the 24th ACM International Conference on Multimedia*. ACM, Amsterdam The Netherlands, 1183–1186. https://doi.org/10.1145/2964284.2973797

[23] Luca Rossetto, Oana Inel, Svenja Lange, Florian Ruosch, Ruijie Wang, and Abraham Bernstein. 2023. Multi-Mode Clustering for Graph-Based Lifelog Retrieval. In *Proceedings of the 6th Annual ACM Lifelog Search Challenge*. ACM, Thessaloniki Greece, 36–40. https://doi.org/10.1145/3592573.3593102

[24] Klaus Schoeffmann. 2023. lifeXplore at the Lifelog Search Challenge 2023. In *Proceedings of the 6th Annual ACM Lifelog Search Challenge*. ACM, Thessaloniki Greece, 53–58. https://doi.org/10.1145/3592573.3593105

[25] Florian Spiess, Ralph Gasser, Silvan Heller, Luca Rossetto, Loris Sauter, Milan Van Zanten, and Heiko Schuldt. 2021. Exploring Intuitive Lifelog Retrieval and Interaction Modes in Virtual Reality with Vitrivr-VR. In *Proceedings of the 4th Annual on Lifelog Search Challenge*. ACM, Taipei Taiwan, 17–22. https://doi.org/10.1145/3463948.3469061

[26] Florian Spiess, Ralph Gasser, Heiko Schuldt, and Luca Rossetto. 2023. The Best of Both Worlds: Lifelog Retrieval with a Desktop-Virtual Reality Hybrid System. In *Proceedings of the 6th Annual ACM Lifelog Search Challenge*. ACM, Thessaloniki Greece, 65–68. https://doi.org/10.1145/3592573.3593107

[27] Florian Spiess and Heiko Schuldt. 2022. Multimodal Interactive Lifelog Retrieval with Vitrivr-VR. In *Proceedings of the 5th Annual on Lifelog Search Challenge*. ACM, Newark NJ USA, 38–42. https://doi.org/10.1145/3512729.3533008

[28] Ly-Duyen Tran, Manh-Duy Nguyen, Duc-Tien Dang-Nguyen, Silvan Heller, Florian Spiess, Jakub Lokoč, Ladislav Peška, Thao-Nhu Nguyen, Omar Shahbaz Khan, Aaron Duane, Björn Þór Jónsson, Luca Rossetto, An-Zi Yen, Ahmed Alateeq, Naushad Alam, Minh-Triet Tran, Graham Healy, Klaus Schoeffmann, and Cathal Gurrin. 2023. Comparing Interactive Retrieval Approaches at the Lifelog Search Challenge 2021. *IEEE Access* 11 (2023), 30982–30995. https://doi.org/10.1109/ACCESS.2023.3248284

[29] Ly-Duyen Tran, Manh-Duy Nguyen, Binh T Nguyen, and Liting Zhou. 2023. Myscéal: a deeper analysis of an interactive lifelog search engine. *Multimedia Tools and Applications* 82, 24 (2023), 37789–37806.

[30] Quang-Linh Tran, Ly-Duyen Tran, Binh T. Nguyen, and Cathal Gurrin. 2023. MemoriEase: An Interactive Lifelog Retrieval System for LSC'23. In *Proceedings of the 6th Annual ACM Lifelog Search Challenge, LSC 2023, Thessaloniki, Greece, June 12-15, 2023*. ACM, 30–35. https://doi.org/10.1145/3592573.3593101