

# **Location-based Queries and Query representation for the Lifelog Search Challenge 2021**

Bachelor thesis

Natural Science Faculty of the University of Basel  
Department of Mathematics and Computer Science  
Databases and Information Systems (DBIS)  
<https://dbis.dmi.unibas.ch/>

Examiner: Prof. Dr. Heiko Schuldt  
Supervisor: Silvan Heller

Sanja Popovic  
[sanja.popovic@stud.unibas.ch](mailto:sanja.popovic@stud.unibas.ch)  
18-054-593

28.07.2021

## **Abstract**

This thesis discusses the thesis template using some examples of the Turing Machine.

# Table of Contents

<b>Abstract</b>	<b>ii</b>
<b>1 Introduction</b>	<b>1</b>
<b>2 Body of the Thesis</b>	<b>2</b>
2.1 Structure . . . . .	2
2.1.1 Sub-Section . . . . .	2
2.1.1.1 Sub-Sub-Section . . . . .	2
2.2 Equations . . . . .	2
2.3 Tables . . . . .	2
2.4 Figures . . . . .	0
2.5 Packages . . . . .	0
<b>3 Conclusion</b>	<b>1</b>
<b>4 Notes</b>	<b>2</b>
4.1 LSC 2020 Papers . . . . .	2
4.1.1 Interactive Lifelog Retrieval with vitrivr (LSC20_DBIS) . . . . .	2
4.1.2 Exquisitor at the Lifelog Search Challenge 2020 . . . . .	3
4.1.3 VRLE: Lifelog Interaction Prototype in Virtual Reality . . . . .	3
4.1.4 Myscéal: An Experimental Interactive Lifelog Retrieval System for LSC'20 . . . . .	4
4.1.5 lifeXplore at the Lifelog Search Challenge 2020 . . . . .	4
4.1.6 A Multi-level Interactive Lifelog Search Engine with User Feedback . .	4
4.1.7 LifeGraph: a Knowledge Graph for Lifelogs . . . . .	5
4.1.8 BIDL-HCMUS@LSC2020: An Interactive Multimodal Lifelog Re- trieval with Query-to-Sample Attention-based Search Engine . . . . .	5
4.1.9 Multimodal Retrieval through Relations between Subjects and Ob- jects in Lifelog Images . . . . .	5
4.1.10 LifeSeeker 2.0 : Interactive Lifelog Search Engine at LSC 2020 . . . .	5
4.1.11 VIRET Tool with Advanced Visual Browsing and Feedback . . . . .	5
4.1.12 SOMHunter for Lifelog Search . . . . .	6
4.1.13 FIRST - Flexible Interactive Retrieval SysTem for Visual Lifelog Ex- ploration at LSC 2020 . . . . .	6

---

4.1.14 Voxento: A Prototype Voice-controlled Interactive Search Engine for Lifelogs . . . . .	6
<b>Appendix A Appendix</b>	<b>7</b>
<b>Declaration on Scientific Integrity</b>	<b>8</b>

# 1

## Introduction

This is the introduction to the thesis template. The goal is to give students a starting point on how to format and style their Bachelor or Master thesis<sup>1</sup>.

Please make sure to always use the most current version of this template, by downloading it always from the original git repository:

<http://www.github.com/ivangiangreco/unibas-latex>

We will use throughout this tutorial some references to Turing's imitation game [?] and the Turing machine [?]. You may be interested in reading these papers.

The package comes with an option regarding the bibliography style. You can include the package with

```
\usepackage[citeauthor]{basilea}
```

to be able to cite authors directly with

```
\citet{turing:1950}
```

If the option is enabled, then the following reference should print Turing [2]: ? ]

---

<sup>1</sup> This document also shows how to use the template.

# 2

## Body of the Thesis

This is the body of the thesis.

### 2.1 Structure

#### 2.1.1 Sub-Section

##### 2.1.1.1 Sub-Sub-Section

#### Paragraph

**Even Sub-Paragraph** This is the body text. Make sure that when you reference anything you use labels and references. When you refer to anything, you normally capitalise the type of object you reference to, e.g. Section 2.1 instead of section 2.1. You may also just use the `cref` command and it will generate the label, e.g., for Section 2.1, we did not specify the word “Section”.

Hint: Try to structure your labels as it is done with `sec:my-label` and `fig:machine`, etc.

### 2.2 Equations

A Turing Machine is a 7-Tuple:

$$M = \langle Q, \Gamma, b, \Sigma, \delta, q_0, F \rangle \quad (2.1)$$

A Turing Machine is a 7-Tuple even if defined in the text, as in  $M = \langle Q, \Gamma, b, \Sigma, \delta, q_0, F \rangle$ .

Table 2.1: Frequency  
after the caption, oth

### 2.3 Tables

Some tables can also be used as shown in Table 2.1<sup>2</sup>. Remember that tables might be positioned elsewhere in the document. You can force positioning by putting a `ht!` in the definition.

Title
The chemical ba
On computable n
Computing mach

<sup>2</sup> Table captions are normally above the table.

## 2.4 Figures

Figures are nice to show concepts visually. For organising well your thesis, put all figures in the Figures folder. Figure ?? shows how to insert an image into your document. Figure 2.2 references a figure with multiple sub-figures, whereas the sub-figures are referenced by Fig. 2.2(a), etc.

## 2.5 Packages

These packages might be helpful for writing your thesis:

**caption** to adjust the look of your captions

**glossaries** for creating glossaries (also list of symbols)

**makeidx** for indexes and the back of your document

**algorithm**, **algorithmicx**, **algpseudocode** for adding algorithms to your document

Turing Machi

(a) Turing Mach

Figure 2.2: Plots of f

Missing: Description  
figure.

# 3

## Conclusion

This is a short conclusion on the thesis template documentation. If you have any comments or suggestions for improving the template, if you find any bugs or problems, please contact me.

Good luck with your thesis!



# 4

## Notes

### 4.1 LSC 2020 Papers

#### 4.1.1 Interactive Lifelog Retrieval with vitrivr (LSC20.DBIS)

what is LSC: In the lifelog search challenge, different multimedia retrieval systems compete in finding specific data. The teams must find a video or a picture as fast as possible. Therefore, it is important to make the queries as efficient and as user friendly as possible.

Dataset: The dataset consists of 115 days of videos and pictures(193'911) collected during a lifelog. Lifelogger record their everyday life (body mounted camera). Thus, the dataset is huge and is perfectly suitable and advantageous for testing, developing and improving multimedia retrieval systems and has spatio-temporal features. To make it possible to query for the images, they are equipped with information (such as local time, geo location, sensor readings for heart rate, tepos, speed and other metadata).

vitrivr: vitrivr is a multimedia retrieval stack and consists of three layers. In the third layer, Cottontail DB, the data is permanently saved. The middle layer, Cineast, is responsible for the extraction of features and processing queries from the first layer, Vitrivr NG. With Vitrivr NG, the user can formulate and combine queries and in the end, get the result data visualized. It is possible to retrieve audio, images, video and 3D models.

*how to get multiple query containers? (red part in paper)*

Existing functionalities in vitrivr:

- Pictures taken in a day belong together such as keyframes in videos. The data model is called *Image Sequence Media Type*. Each single images of the serie is mapped to a media segment. The media segments belong to a single media object in the data model. (The features are generated per media segment)
- With boolean queries, the user can search for a specific time, location, time span and so on.(example weekday=Wednesday)
- With Vitrivr NG, it is possible to refine the result set with filters during browsing.

New functionalities:

- temporal scoring-it is possible to add a time property where the user can query "canyon occurs before bridge **within 30s.**". After the queries for "bridge" and "canyon" are executed and fused, the multimedia objects of the result set are then examined (for the time). The segments of the multimedia objects are temporally ordered.
- deep learning based semantic concepts
- staged querying: Previously, when querying for a building and a white blob, two lookups would run independently against the entire collection. There would be only a few results and a lot of false positives only containing one of the required features because of the OR relation. With the new early filtering, the user can first query for a building, then on the new resulting subset instead of the entire collection, query for white blob. With this technique, there are less false negatives.

#### 4.1.2 Exquisitor at the Lifelog Search Challenge 2020

Exquisitor is an interactive learning system. With the help of user, the system "learns" what the user wants. The user helps with deciding which images of a suggested collection match the description on a task in LSC. The user either accepts or declines an image. If an image is accepted, the system suggests a new image which resembles (has similar features as) the accepted image.

The classification takes place with the help of support vector machine (SVM), a machine learning algorithm. The system learns what the user wants. Through the interactive learning interface, the system initially suggests a collection of images from the LSC dataset and the user gives feedback on those (accept or decline). With the help of the feedback from the user, the system learns, and provides new suggestions (iterates as long as user isn't satisfied).

#### 4.1.3 VRLE: Lifelog Interaction Prototype in Virtual Reality

This project focusses more on the interaction methods between user and the VR-device. The VR-device offers two tabs: in one tab, the user can choose the temporal aspect of a query (day, time) and in another tab, the user can choose zero to many concepts such as apple, airport, altar, area etc. The concepts are located in subtabs (alphabetically ordered). There are two developed interaction methods in this project. In the first method, the user interacts with the UI via controllers equipped with beams (Distance-Based Interaction). In the second method, the user touches the screen with drumstick-like sticks (Contacted-Based Interaction).

Initially, features as text, times and GPS coordinates etc. from the images of the dataset are extracted and indexed.

It is discussed how the user can find an image as fast as possible. Since the VRLE system works with events (multiple images in one event) and since the LSC task is about finding a specific image, an event is presented with nine ranked keyframes. The ranking is done by the search engine. If for example all images would have been presented, the user might be overwhelmed. The developers engaged in optimizing the user experience by questioning

which keyframes should be shown and how should the user interact with the system (query formulation and actual interaction).

#### 4.1.4 Myscéal: An Experimental Interactive Lifelog Retrieval System for LSC'20

The Myscéal system supports natural language queries. The user queries for example "having coffee at the helix at 9 am on 23 August" (taken from paper)(the system will recognize that there is an object (coffee), location (the helix), time (9 am), and date (23 August)). The system then extracts information such as location, action, time and date. Sequential queries are also supported. The user can provide information about the actions before and after the sought event. Mysceal has the same search engine as "VRLE: Lifelog Interaction Prototype in Virtual Reality". In general, information such as location, time, activities and visual concepts are saved. GPS information from the metadata are saved as longitude and latitude. Besides the mentioned information, an object detector is used. With object detection, the system can also evaluate how relevant an object is in the image. **An important addition is the possibility to query for times GPS data. The user can draw a rectangle on the map.**

#### 4.1.5 lifeXplore at the Lifelog Search Challenge 2020

Images from a day are bundled and converted into a video.

LifeXplore offers various *feature maps* organized by for example "car" or "person". A map is big and consists of keyframes. It would be too time-consuming for a LSC task, but in general, the user could explore the map and could do further query refinement.

A *day summary view* is offered, where the user can look at all images of a specific day.

LifeXplore offers multiple retrieval options: concept and metadata search, sketch search as well as similarity search(taken from paper). The so called *Filter View* offers combining multiple filters. (activate the weekday filter for only retrieving weekdays together with a time range filter for "10:00-14:30" in combination with a named location from the metadata, "The Helix". The result view on the right-hand side contains all retrieved shots ordered by video. Besides the most basic filters such as weekday and time, lifeXplore includes following additional filters: deep concepts, speed, heart rate, number of steps, calories burned, activity type and geographic location-taken from paper)

With the *Sketch Search View*, the user can sketch and hence retrieve keyframes (Figure 5 in paper). The system also provides a similarity search (lifeXplore tries to find similar shots according to the current user-chosen similarity measure-paper).

#### 4.1.6 A Multi-level Interactive Lifelog Search Engine with User Feedback

While preprocessing the data, blurry images and duplicates are removed and features extracted (three level feature extraction). Three level feature extraction: visual, textual,

behavior features.

The user can provide temporal aspects to query. Possibilities: text as query ((similar to Exquisitor) the user can give negative feedback on images and thus exclude them from the result list and then the system can get more information from the user that cannot be extracted from the queries), image as query (user can upload image and similar images are suggested), timeline view.

#### 4.1.7 LifeGraph: a Knowledge Graph for Lifelogs

#### 4.1.8 BIDAL-HCMUS@LSC2020: An Interactive Multimodal Lifelog Retrieval with Query-to-Sample Attention-based Search Engine

The user can create text queries and query with images. Filters are also available, so that the user can enrich the query. The user can modify the result set by removing inaccurate images.

#### 4.1.9 Multimodal Retrieval through Relations between Subjects and Objects in Lifelog Images

With this system, the user can formulate text queries or provide keywords such as time, location, activity and concepts. Datetime and location can also be specified by the user. **A map with in the lifelog visited locations is shown.**

#### 4.1.10 LifeSeeker 2.0 : Interactive Lifelog Search Engine at LSC 2020

**This system has a different approach for visualizing location aspects-interactive transitional graph based filter. Vertices in the graph represent locations and the directed edges represent the chronological order. The graph can be represented on three different levels: country, location, area of location. If the user travelled from Switzerland to Germany, then on the country level, a directed edge would go from vertex Switzerland to vertex Germany. The user can click on a vertex and the system would then display relevant images taken there.**

When the user clicks on an image chosen from the result set, the chosen image appears in a pop-up together with images taken before and after. In general, a user can create free-text queries and also filter with time and location keywords etc. or use the graph-based filter.

The developers also have taken care of extracting textual information (text in images). With this, it is possible to identify what the lifelogger is doing (activity).

#### 4.1.11 VIRET Tool with Advanced Visual Browsing and Feedback

The updates in 2020 of the VIRET system focussed on improving the result set visualization. *Images of a day - bundled together (as keyframes, one video).*

It is possible, to create free-text queries and to apply filters (date, biometrics, GPS). Further-

more, the user can create temporal queries by describing two consecutive shots. New improvement is to better visualize ranked images. (Moreover, the target scene frame (or visually similar one) could get promising rank with respect to the overall collection size (e.g., rank 1000 out of one million), yet the frame is practically unreachable with the sequential browsing approach and a time limit to solve the task. To address these issues, our primary objective focuses on the presentation of top ranked results of a provided query.-paper) To help the user to understand the result set, descriptors are added to image, which the user can see.

#### 4.1.12 SOMHunter for Lifelog Search

Procedure: The user enters a text query (also possible to add days, time) and the system returns a map of images. As an intermediate step, the user then selects images that seem relevant and the system proposes new images (a new map) with temporal context. With temporal context they mean displaying top ranked images with according frames that were taken before and after (Figure 1-paper).

The user can access previous maps, if the new generated map doesn't have relevant images.

#### 4.1.13 FIRST - Flexible Interactive Retrieval SysTem for Visual Lifelog Exploration at LSC 2020

entering query in text format

System extracts text from image such as brand name, shop names, street names. Also, the team manually added captions to some images and later, the system's "captioning module" could then caption the whole lifelog dataset. The system also extracts personalized concepts of the lifelogger (his everyday visual objects that are not available outside). The system also can group consecutive images, provides querying by example, querying by text. The user can mark images as negative example and the system eliminates images that are similar to it.

#### 4.1.14 Voxento: A Prototype Voice-controlled Interactive Search Engine for Lifelogs

The user creates, deletes and ends queries by speaking. The user can manually edit the text which the system extracted from the users spoken sentence. The user can speak continuously and *see in real-time the updates to the result set as they continue to formulate a long query.-paper* The response of the system is also spoken (and visualized).



## **Appendix**

# Declaration on Scientific Integrity

## Erklärung zur wissenschaftlichen Redlichkeit

includes Declaration on Plagiarism and Fraud  
beinhaltet Erklärung zu Plagiat und Betrug

**Author — Autor**

Sanja Popovic

**Matriculation number — Matrikelnummer**

18-054-593

**Title of work — Titel der Arbeit**

Location-based Queries and Query representation for the Lifelog Search Challenge 2021

**Type of work — Typ der Arbeit**

Bachelor thesis

**Declaration — Erklärung**

I hereby declare that this submission is my own work and that I have fully acknowledged the assistance received in completing this work and that it contains no material that has not been formally acknowledged. I have mentioned all source materials used and have cited these in accordance with recognised scientific rules.

Hiermit erkläre ich, dass mir bei der Abfassung dieser Arbeit nur die darin angegebene Hilfe zuteil wurde und dass ich sie nur mit den in der Arbeit angegebenen Hilfsmitteln verfasst habe. Ich habe sämtliche verwendeten Quellen erwähnt und gemäss anerkannten wissenschaftlichen Regeln zitiert.

Basel, 28.07.2021

---

**Signature — Unterschrift**