



Funded by
the European Union

SUPPORTED
BY



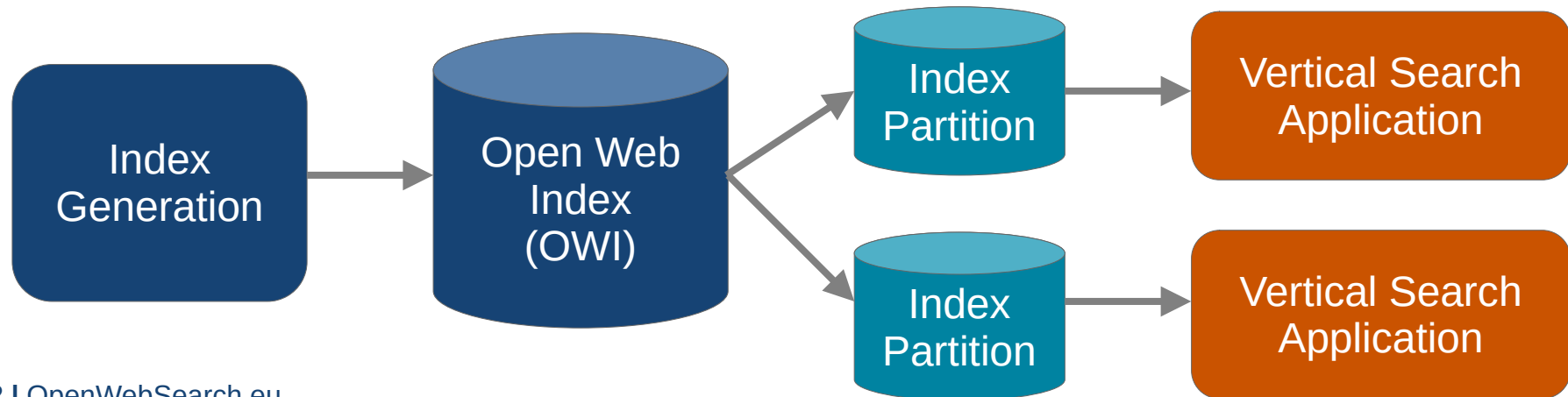
Conceptual Design and Implementation of a Prototype Search Application using the Open Web Index

Alexander Nussbaumer, Rohit Kaushik,
Gijs Hendriksen, Sebastian Gürtl, Christian Gütl

5th International Open Search Symposium (OSSYM)
4-6 October 2023

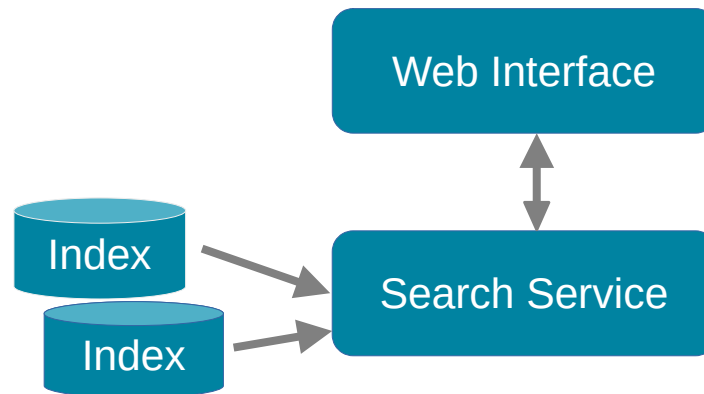
Vertical Search Engines in OWS

- Vertical Search Engines serve special purposes, topics, or domains, e.g. product search
- Search applications are a key concept in OpenWebSearch.eu
- Based on Open Web Index and index partitions



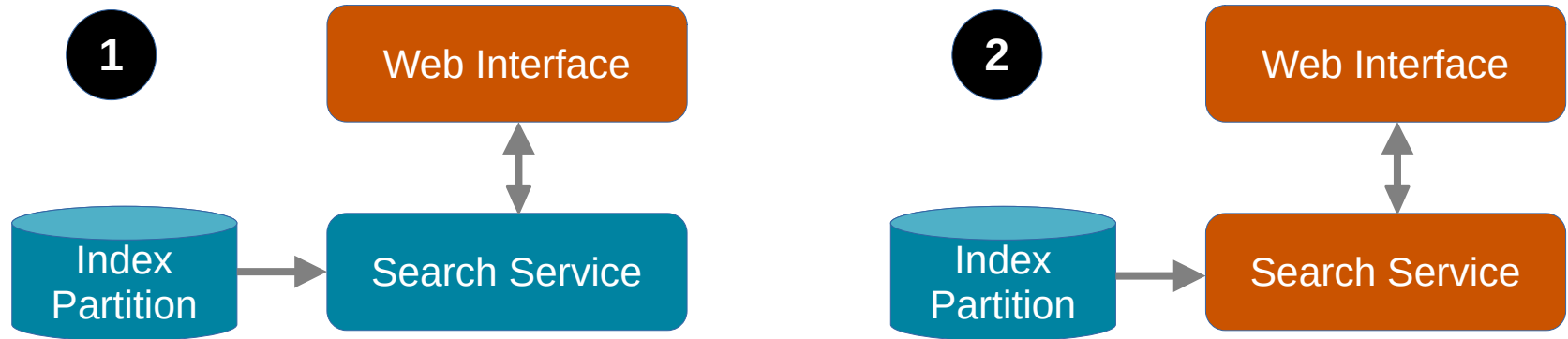
Prototype Search Application

- Minimal implementation of a OWS vertical search engine
 - Index partitions in reasonable size on a local server
 - Search service that searches the index and provides REST API
 - Web interface that allows end-users specific search

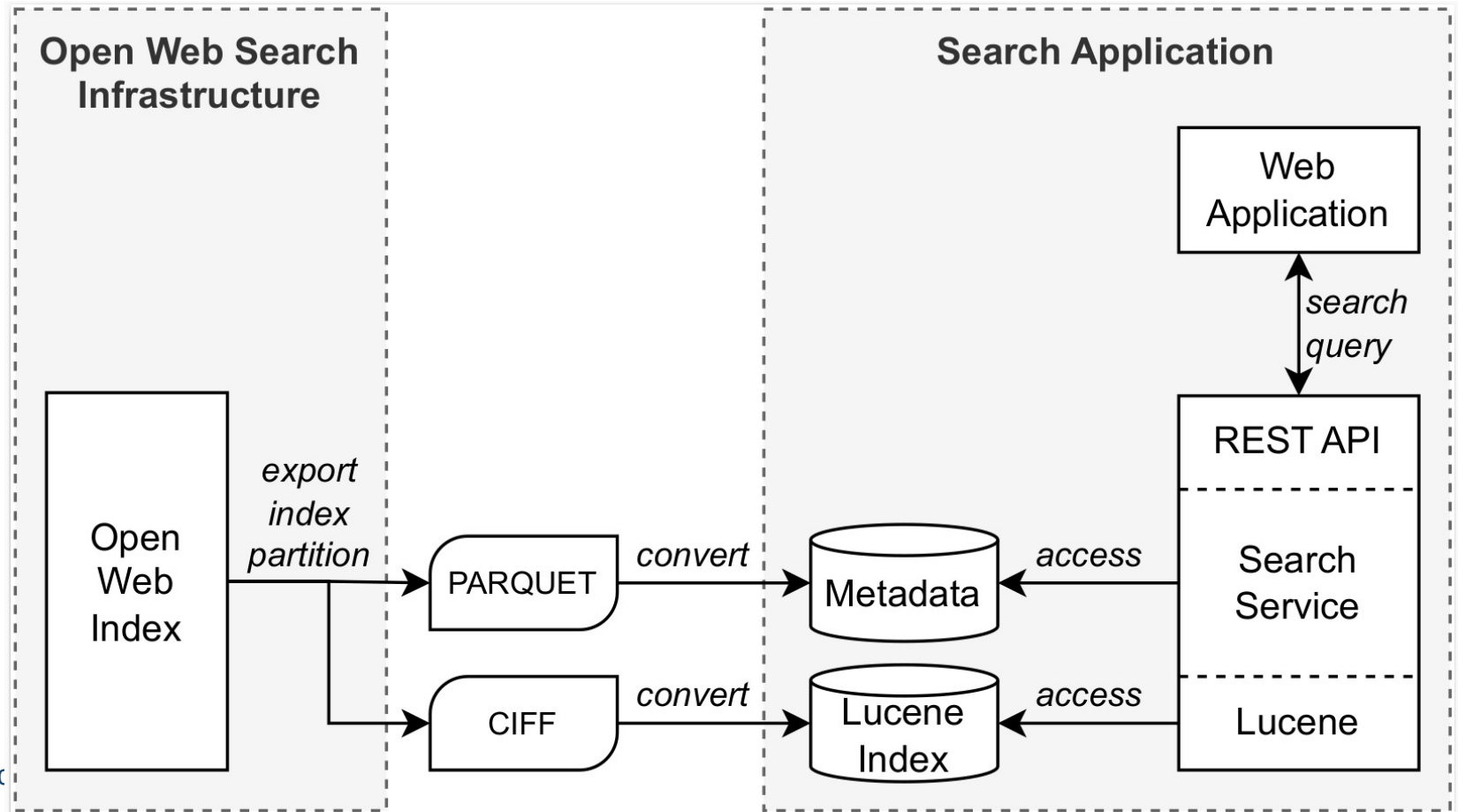


Prototype Search Application

- Guides the development of a OWS search application
 - Using the search service and development of a web interface (1)
 - Using the concept and development of the whole application (2)



Software Architecture



Index Partition: CIFF and PARQUET

- CIFF: Index exchange format; inverted index
- Parquet: Tabular format

CIFF

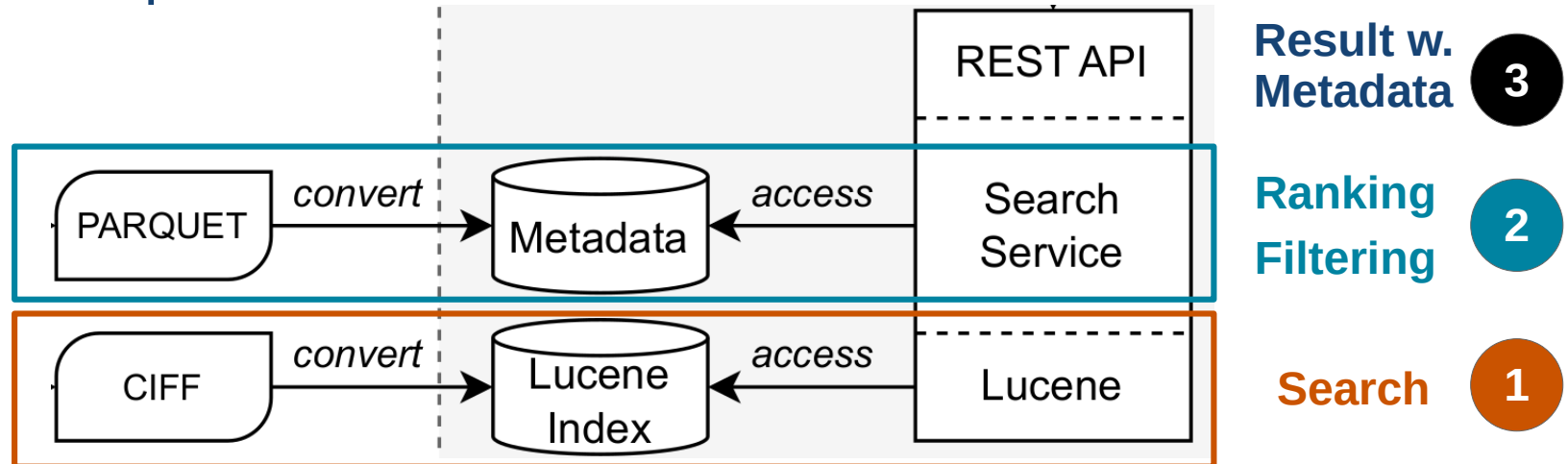
Term	List of (Document, Frequency)
Term 1	(docID A, 1) (docID B, 3) (docID C, 1)
Term 2	(docID D, 2)
Term 3	(docID E, 4) (docID F, 2)
Term 4	(docID H, 2) (docID I, 1)

Parquet

Document	Lang	Full text	WARC date	...
docID A	en	... text ...	2023-07-01	
docID B	de	... text ...	2023-07-01	
docID C	de	... text ...	2023-07-01	
docID D	en	... text ...	2023-07-01	

Search Process

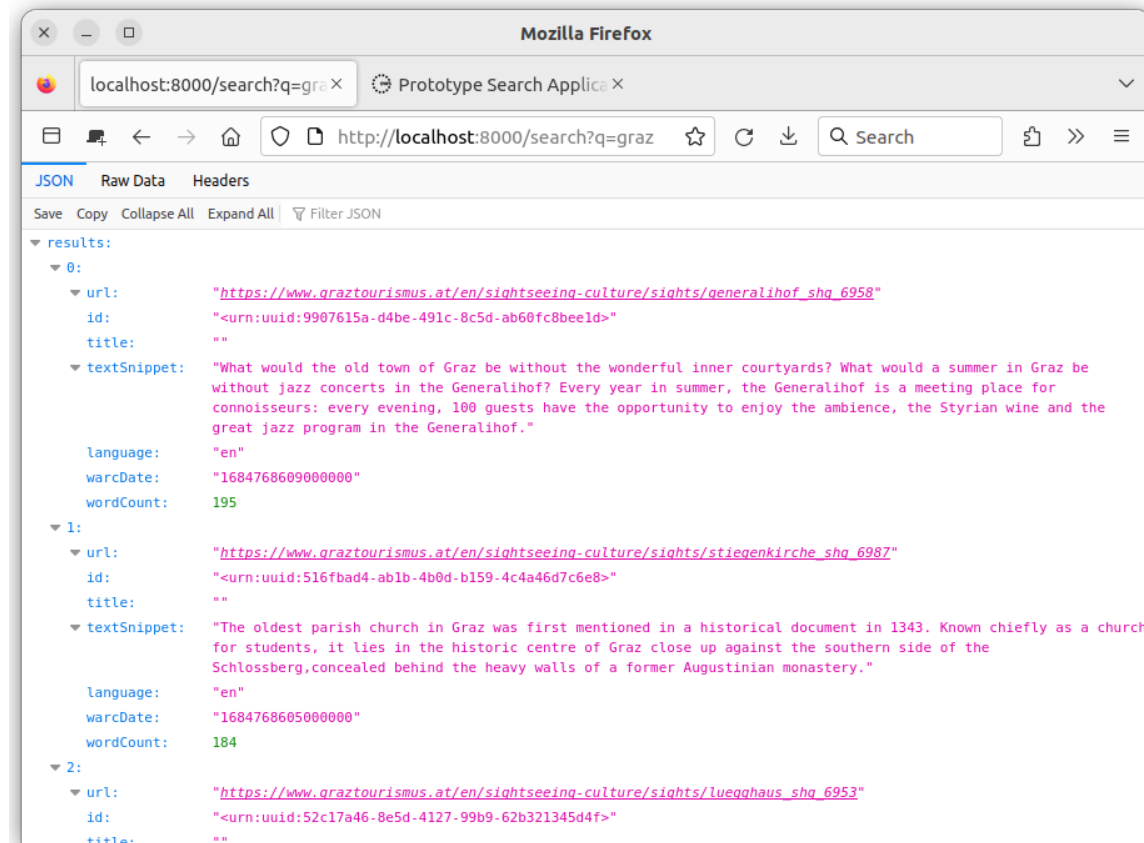
- CIFF file is converted to a Lucene index that can be searched with the Lucene library
- Search result is filtered and ranked using metadata coming from the Parquet file



REST API

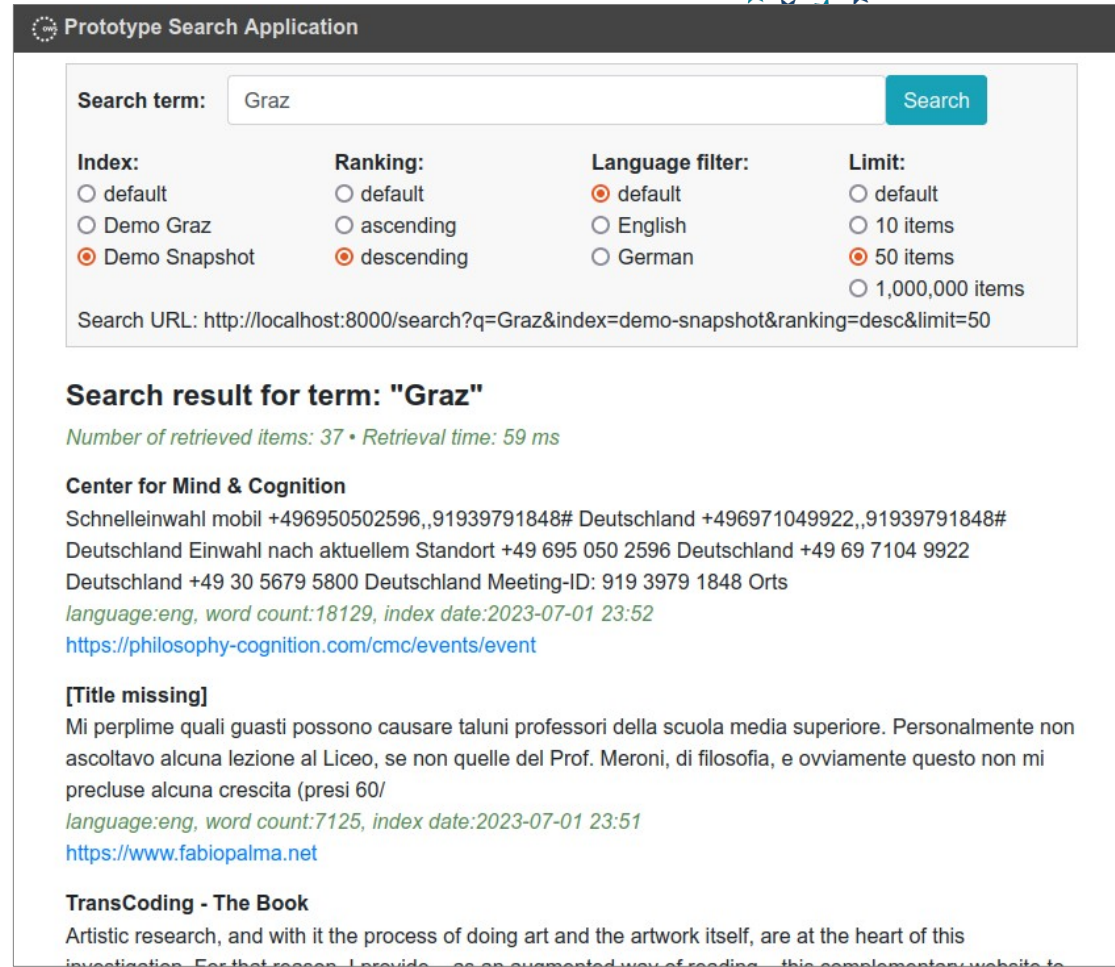


- End point:
 - host:port/search?...
- Request parameters
 - Search term, Index, ranking, language, limit
- Response data
 - URL, title, text snippet, WARC date, word count



Demonstration 1: Basic Search

- Simple web interface to the Search Service
- Operates two indices
- Performs ranking and filtering
- Displays metadata



Prototype Search Application

Search term:

Index:
 default
 Demo Graz
 Demo Snapshot

Ranking:
 default
 ascending
 descending

Language filter:
 default
 English
 German

Limit:
 default
 10 items
 50 items
 1,000,000 items

Search URL: `http://localhost:8000/search?q=Graz&index=demo-snapshot&ranking=desc&limit=50`

Search result for term: "Graz"
Number of retrieved items: 37 • Retrieval time: 59 ms

Center for Mind & Cognition
Schnelleinwahl mobil +496950502596,,91939791848# Deutschland +496971049922,,91939791848#
Deutschland Einwahl nach aktuellem Standort +49 695 050 2596 Deutschland +49 69 7104 9922
Deutschland +49 30 5679 5800 Deutschland Meeting-ID: 919 3979 1848 Orts
language:eng, word count:18129, index date:2023-07-01 23:52
<https://philosophy-cognition.com/cmc/events/event>

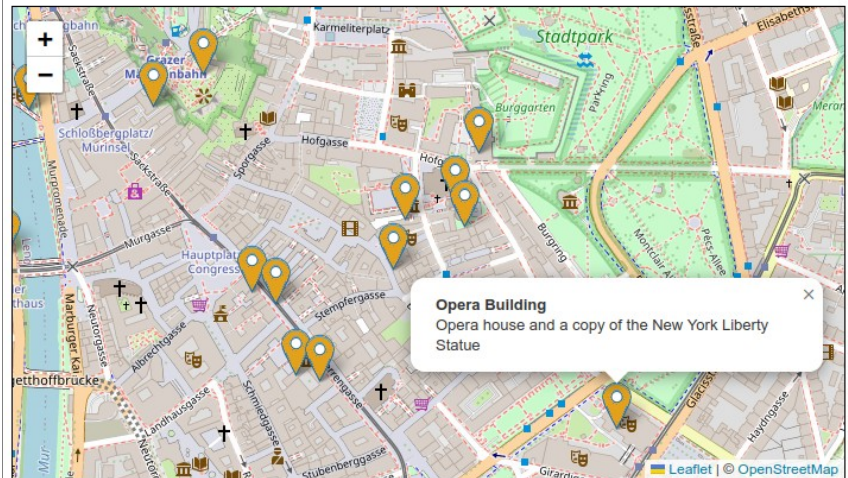
[Title missing]
Mi perlime quali guasti possono causare taluni professori della scuola media superiore. Personalmente non ascoltavo alcuna lezione al Liceo, se non quelle del Prof. Meroni, di filosofia, e ovviamente questo non mi preclude alcuna crescita (presi 60/
language:eng, word count:7125, index date:2023-07-01 23:51
<https://www.fabiopalma.net>

TransCoding - The Book
Artistic research, and with it the process of doing art and the artwork itself, are at the heart of this investigation. For that reason, I provide – as an augmented way of reading – this complementary website to

Demonstration 2: Sightseeing Application

- Integration into existing Sightseeing application Spots
- Clicking on a sightseeing marker triggers a search with the title as search term

Sightseeing Graz

A map of Graz, Austria, with several orange location markers. A pop-up window is open over one marker, displaying search results for "Opera Building". The map includes street names like "Hofgasse" and "Stemplergasse", and landmarks like "Stadtspark" and "Karmeliterplatz".

Opera Building
Opera house and a copy of the New York Liberty Statue

Search result for term: Opera Building

Best attractions in Graz
The cafe, opened in a hotel with the same name, can be called a full-fledged Austrian attraction, because it is here that they serve the real Sacher cake, cooked according to the original ancient recipe. Keep in mind that there are always a lot of vi
<https://www.tripzaza.com/en/destinations/top-attractions-in-graz>

Opera House Graz
Something fascinating about Graz - the interplay of modernism and tradition - is illustrated by the sculpture "light sword" next to the opera house. It was originally made for the festival "steirischer herbst" in 1992. To celebrate the 500th annivers
https://www.graztourismus.at/en/sightseeing-culture/sights/opera-house_shg_1472

Graz Opera
Past general music directors (GMD) of the company have included Niksa Bareza (1981-1990), Philippe Jordan (2001-2004), Johannes Eritzsch (2006-2013), and Dirk Kaftan (2013-2017) [7] In

Create You Own Search Application!



- Code available on GitLab in public area
- Follow instructions in Readme file to download and start service
- Try out REST API in your browser
- Try out Web interface in your web server
- Use the code to create you own application

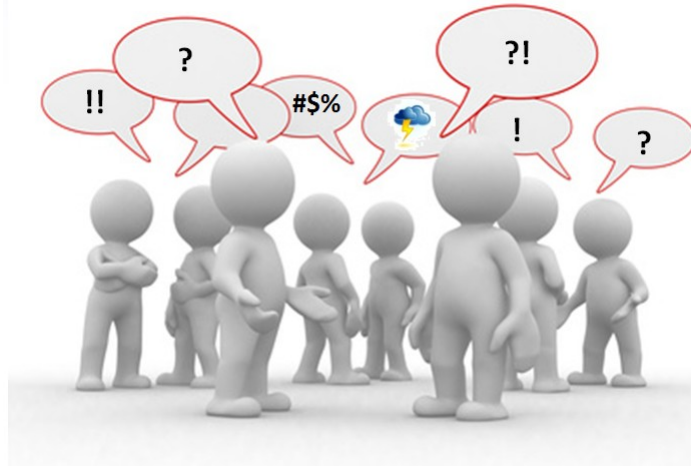


<https://opencode.it4i.eu/openwebsearcheu-public/prototype-search-application>

Next Steps

- CIFF and Parquet file Management
 - Alignment CIFF and Parquet files with OWS infrastructure
 - Merging multiple CIFF files and related Parquet files
 - Performance tests with large index partitions
- Development environment for creation search application
 - Depending on feature requests from Hackathons, OWS Applications, 3rd party applications, etc.
- Ethical and legal implications

Any Questions?



Contact:

- Alexander Nussbaumer <alexander.nussbaumer@tugraz.at>
- Sebastian Gürtl <sguertl@student.tugraz.at>
- Graz University of Technology, Austria