TECHNISCHE UNIVERSITÄT DRESDEN

DRESDEN concept

Faculty of Environmental Sciences
Chair of Geoinformatics

LLMs for Conversational Geodata Search

AGILE 2025 Tutorial, Dresden, Germany

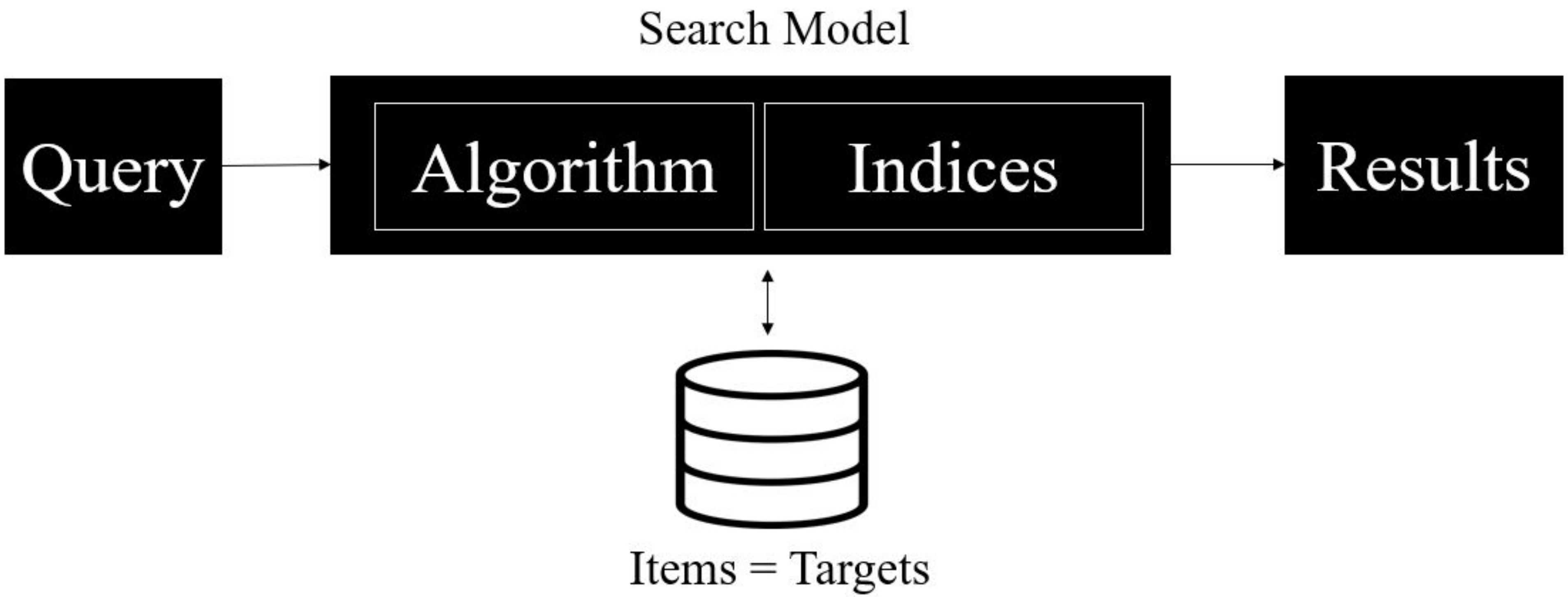
Auriol Degbelo, Simeon Wetzel, Stephan Maes

Agenda

- Search Targets
- GeoQuery
- Retrieval-Augmented Generation

Automated Information Retrieval

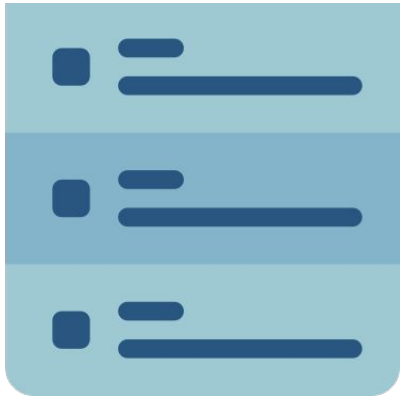
Key abstract components



What are QA Systems?

“Question Answering (QA) Systems can be considered as an extension of Search Engines in the sense, that they aim at automatically **supplying users with precise answers to questions** posed in natural language, instead of simply returning a ranked list of relevant sources based on a set of keywords”. (Dimitrakis et al, 2020)

Where is Dresden located? → Old Search Engines

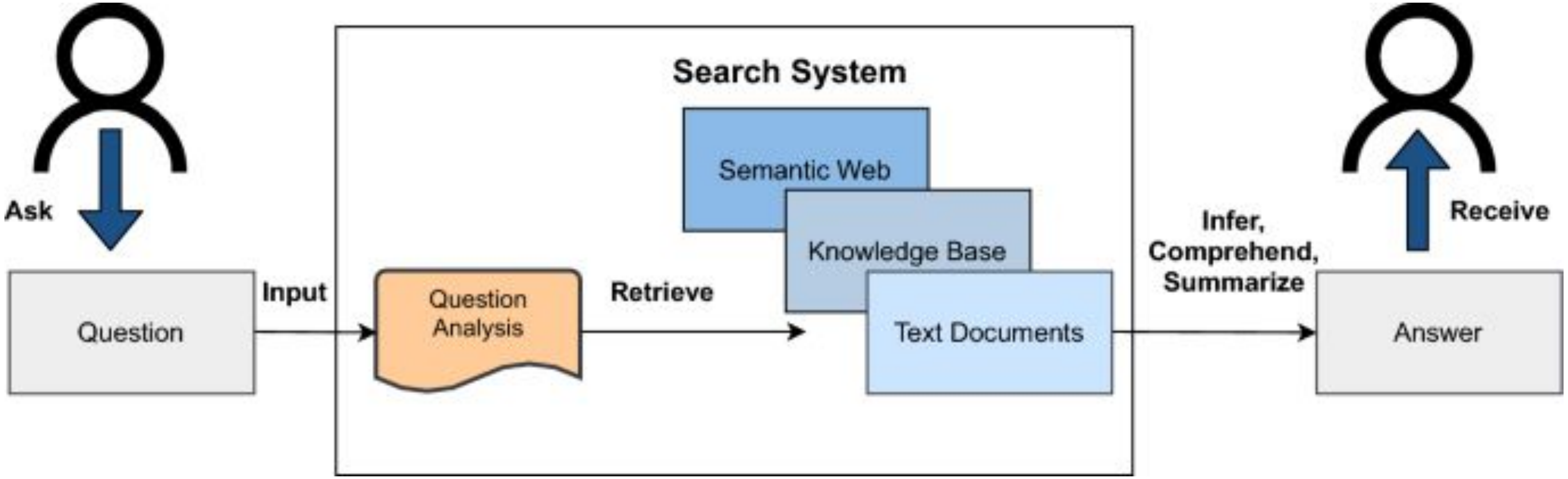


Where is Dresden located? → Q&A System

Dresden is located in Germany ← Q&A System

Question Answering Systems

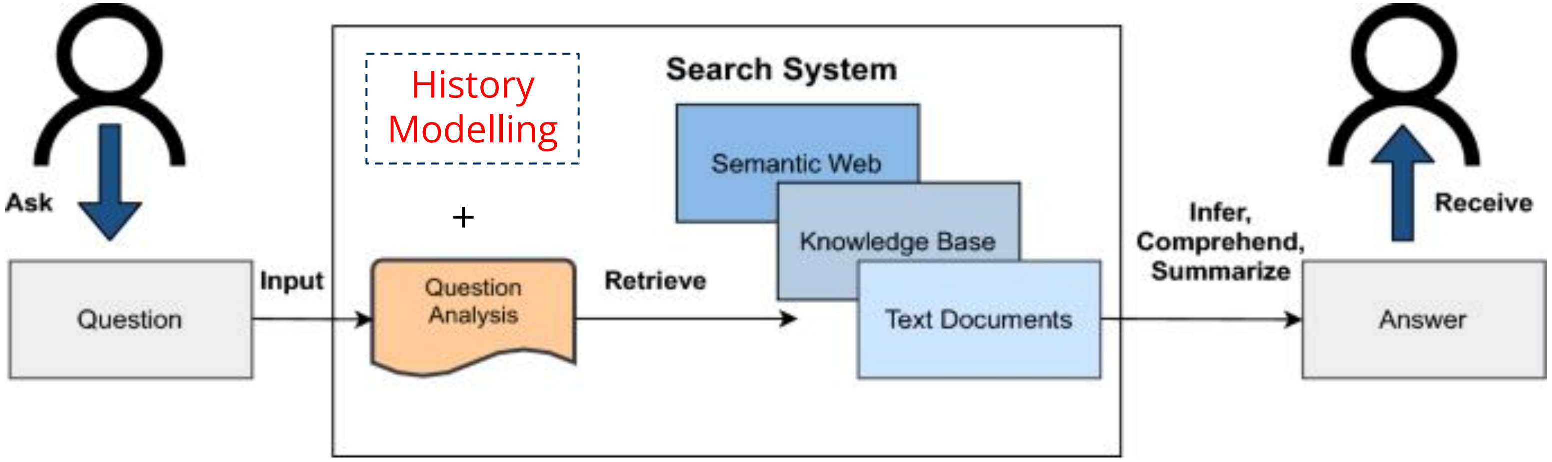
Key abstract components



[Zaib et al, 2022]

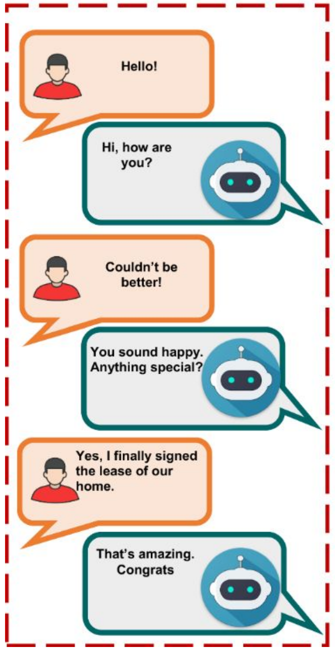
Conversational Question Answering Systems

Key abstract components

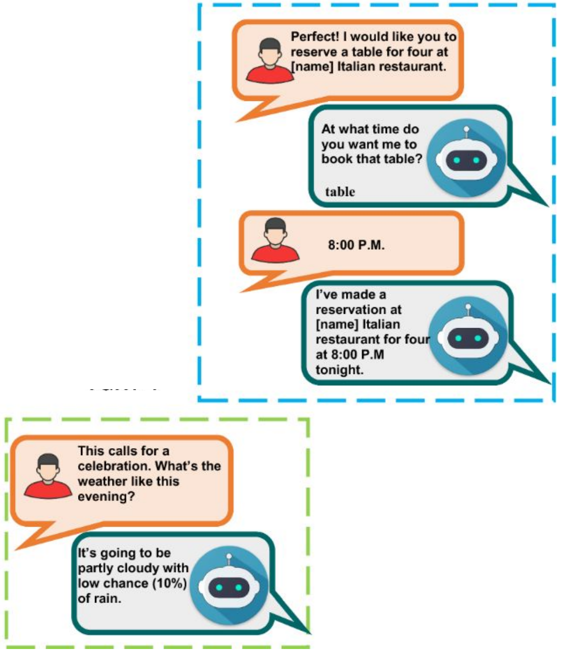


[Zaib et al, 2022]

Chat-Oriented



Task-Oriented



TECHNISCHE UNIVERSITÄT DRESDEN

Search Targets, GeoQueries

Search Target

A search target is a resource searched for.

- Georeferenced datasets
- Georeferenced videos
- Raster maps
- Interactive maps
- Geospatial services (e.g. OGC web services)
- Shapefiles
- Algorithms in scientific papers
- …

Geographic Information Retrieval: Text
Multimedia Information Retrieval: Audio, Video, …
Dataset Search: Dataset
…

Search target for our scenario: document

metadata

```
{ 'tourism': 'museum',
 'building': 'yes',
 'addr:postcode': '01069',
 'blind:description:de': 'Führunge für Sehgeschädigte und Blinde',
 'fee': 'yes',
 'email': 'service@dhmd.de',
 'addr:housenumber': '1',
 'id': "('relation', 5651)",
 'addr:street': 'Lingnerplatz',
 'heritage': 'yes',
 'dog': 'no',
 'check_date': '2024-03-09',
 'website': 'https://www.dhmd.de',
 'internet_access': 'wlan',
 'name:en': 'German Hygiene Museum',
 'wikidata': 'Q874373',
 'roof:shape': 'flat',
 'roof:colour': '#727466',
 'alt_name:de': 'Hygienemuseum',
 'toilets': 'yes',
 'toilets:wheelchair': 'yes',
 'internet_access:fee': 'no',
 'building:architecture': 'neoclassicism',
 'type': 'multipolygon',
 'wikipedia': 'de:Deutsches Hygiene-Museum',
 'architect': 'Wilhelm Kreis',
 'changing_table': 'yes',
 'building_description': 'yes',
 'wheelchair': 'yes',
 'opening_hours': 'Tu-Su 10:00-18:00',
 'name': 'Deutsches Hygiene-Museum',
 'geometry': '{"type": "Polygon", "coordinates": [[[13.7455854, 51.0450922], [13.7459654, 51.0449423],
 'internet_access:access': 'customers',
 'check_date:opening_hours': '2022-06-20',
 'start_date': '1928..1930',
 'smoking': 'no',
 'addr:city': 'Dresden',
 'phone': '+49 351 4846400'}
```

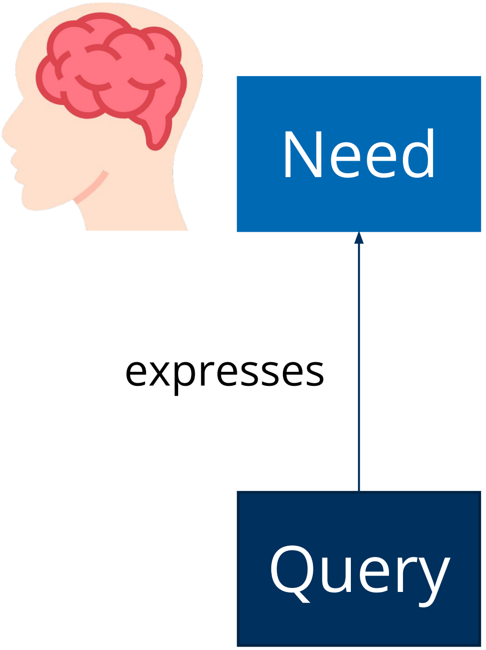
document = id + metadata

Example: id: ('relation', 5651)

Visualization

GeoQuery

Need vs Query



[Baeza-Yates and Ribeiro-Neto, 1999]

A clear statement in natural language that describes what the user is asking for or wants to know.

'I am looking for restaurants near Dresden'.

The expression of the user information need in the <u>input language</u> provided by the information system.

Q1: „restaurant near Dresden"

Q2: SELECT r where ....

Q3

GeoQuery

Definition

Geoquery = A query that requests location-specific information

= The expression of the user’s need for location-specific information in the input language provided by the information system

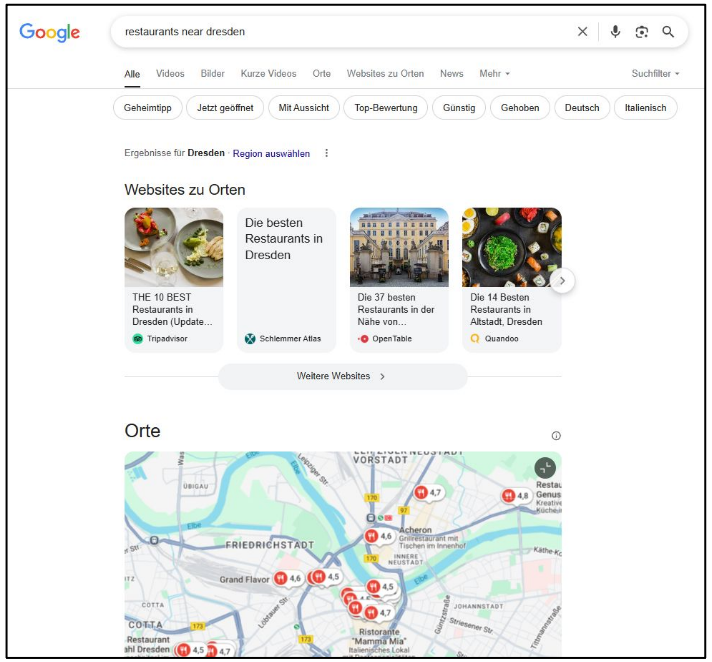
GeoQuery

Two forms of geoqueries

<theme><relationship><space> [Purves et al, 2018]

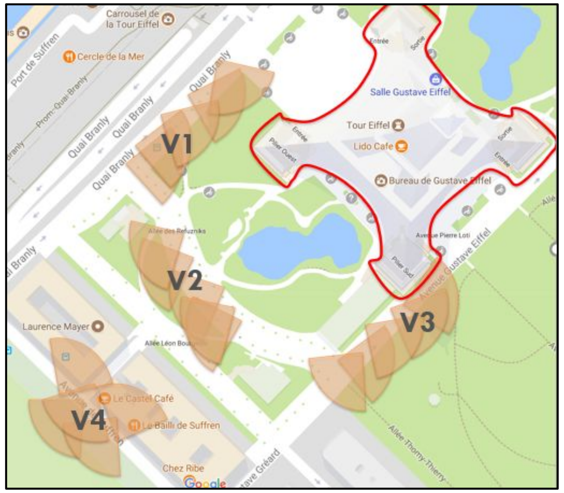
(very often <object><relationship><place> [Carniel, 2023])

- Weather *in* Dresden
- Restaurants *near* Dresden
- Castles *inside* Dresden
- Houses *north of* Dresden



<resources> about <theme><space><time> [Degbelo, 2022]

- Maps about the cities of Prussia in 1830
- Videos about the Eiffel Tower in 2020, whose spatial locations are within *bbox*



GeoQuery Interpretation

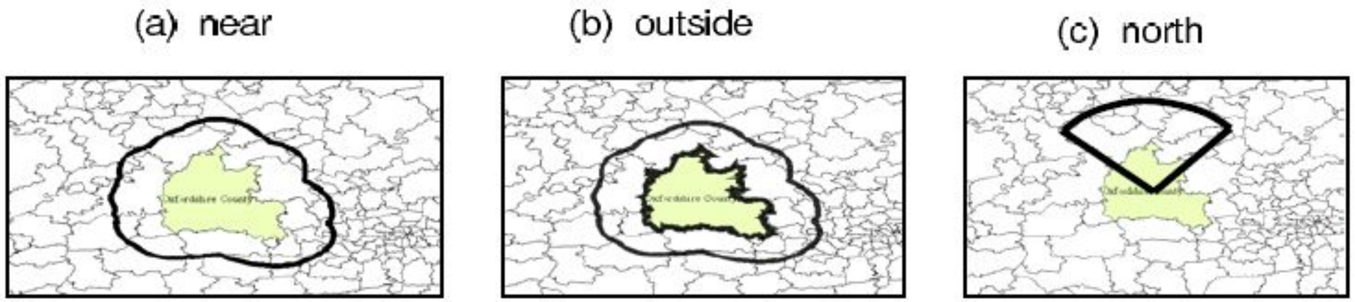
A geoquery has a footprint

The footprint of a spatial term P-footprint indicates the geographical location of the intended place, and is specified in terms of map coordinates with a selected reference system. [Fu et al, 2005]

A query footprint Q-footprint defines a geographical space that covers the intended spatial search extent of the query, and it is specified in the form of map coordinates. [Fu et al, 2005]

Castles ***rel*** Oxfordshire

[Fu et al, 2005]



Retrieval-Augmented Generation

Retrieval-Augmented Generation

“RAG is a technique for augmenting LLM knowledge with additional data.

LLMs can reason about wide-ranging topics, but their knowledge is limited to the public data up to a specific point in time that they were trained on. If you want to build AI applications that can reason about private data or data introduced after a model's cutoff date, you need to augment the knowledge of the model with the specific information it needs. The process of bringing the appropriate information and inserting it into the model prompt is known as Retrieval Augmented Generation (RAG)”. [source]

RAG = LLMs + Your Own Data [CSV, JSON, Webpages, PDF, Graph Data, …]

GeoRAG = LLMs + Your Own Data + Spatially-Explicit Reranking

Retrieval-Augmented Generation

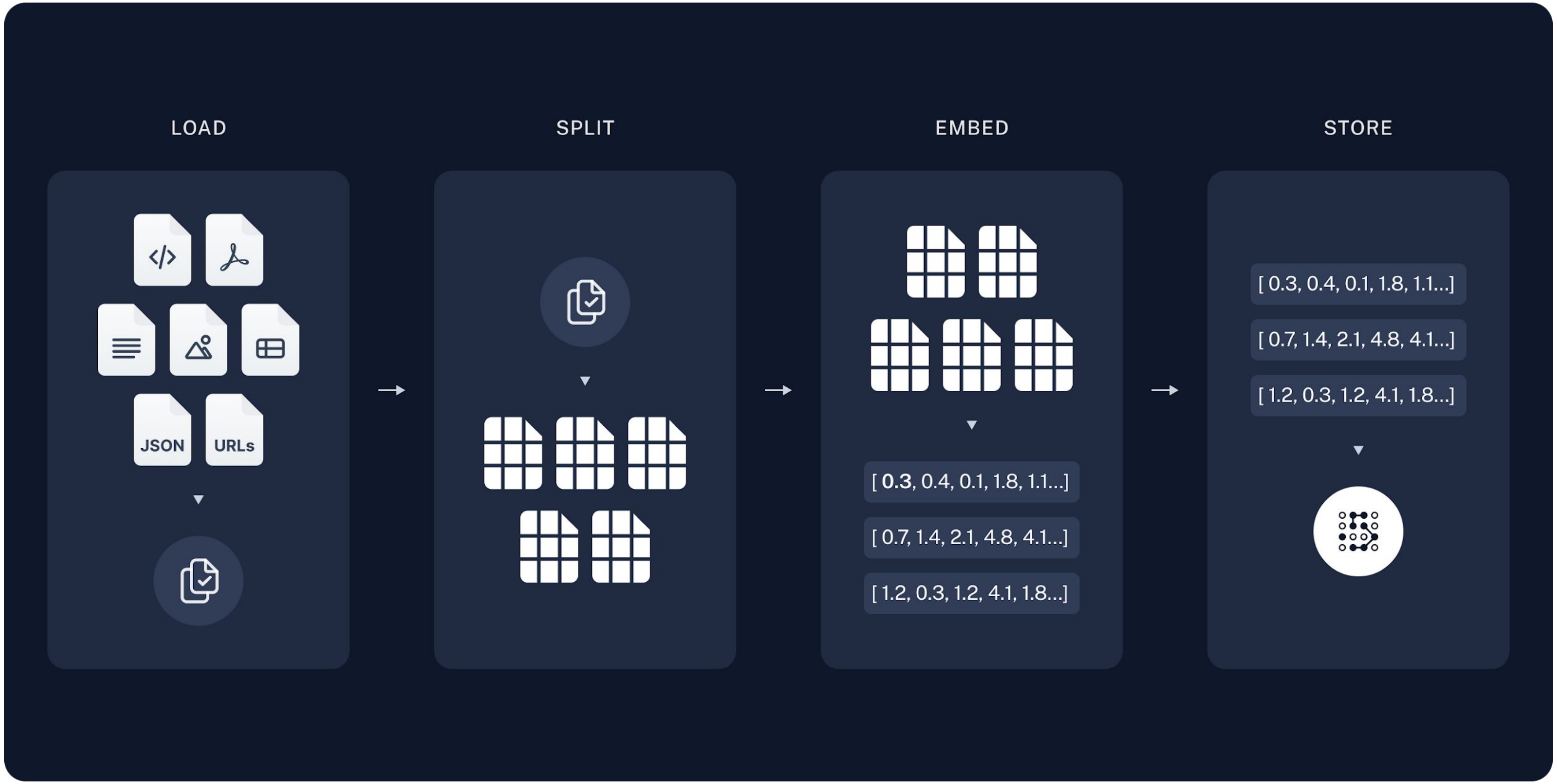
A RAG Application typically needs three steps…

Indexing (dividing targets into smaller chunks and storing these, usually before the user query)

Retrieval (selection of relevant chunks to the user query)

Generation (answer formulation based on the chunks from the previous step)

RAG: Indexing and Storage



https://python.langchain.com/docs/tutorials/rag/

AGILE: GIScience Series, 6, 1, 2025. https://doi.org/10.5194/agile-giss-6-1-2025
Proceedings of the 28th AGILE Conference on Geographic Information Science, 10–13 June 2025.
Eds.: Auriol Degbelo, Serena Coetzee, Carsten Keßler, Monika Sester, Sabine Timpf, Lars Bernard

© Author(s) 2025. This work is distributed under the Creative Commons Attribution 4.0 License.

AGILE

Preface

Auriol Degbelo [1], Serena Coetzee [2], Carsten Keßler [3,4], Monika Sester [5], and Sabine Timpf [6]

[1] Chair of Geoinformatics, TU Dresden, Dresden, Germany
[2] UNU-FLORES, Dresden, Germany
[3] Bochum University of Applied Sciences, Bochum, Germany
[4] Aalborg University, Copenhagen, Denmark
[5] Leibniz University Hannover, Hannover, Germany
[6] University of Augsburg, Augsburg, Germany

We are excited to welcome you to Dresden for the 28th AGILE international conference! This year's conference theme is 'Geographic Information Science responding to Global Challenges' to emphasise the critical role of geospatial data and analysis in informing decision-making for current global issues affecting humanity.

We are pleased to host researchers, developers, educators, students, and professionals from all areas of Geographic Information Science and Technology. AGILE 2025 offers a platform to learn about the latest research in Geographic Information Science, showcase products, and network with colleagues from Europe and the rest of the world.

1 About the host

TU Dresden was founded in 1828 and has been designated a University of Excellence by the German government since 2012. It is the largest university in the federal state of Saxony and one of the largest German universities of technology. It is home to 29,000 students, about 5,800 (20%) of whom are international students from 128 countries. The university offers 119 degree programmes, 26 of which are taught in English and 26 of which are double degree programmes. As of 2025, the university has 8,500 employees, 3,500 of whom are funded by third parties. TU Dresden comprises five schools and 17 faculties.

The Faculty of Environmental Sciences unites the three disciplines of forest, geo-, and hydro sciences, which is unique in Germany. Activities focus on monitoring and modeling the Earth System at global, regional, and local scales, and the use of this knowledge in the context of the sustainable development of the human environment. Research and teaching cover a wide range of topics and are firmly embedded in both regional and international networks. The faculty's commitment to advancing education and research in geographic information science is reflected through the following programmes: BSc Geodesy and Geoinformation, BSc Environmental Informatics, MSc Cartography, and MSc Geoinformation Technologies.

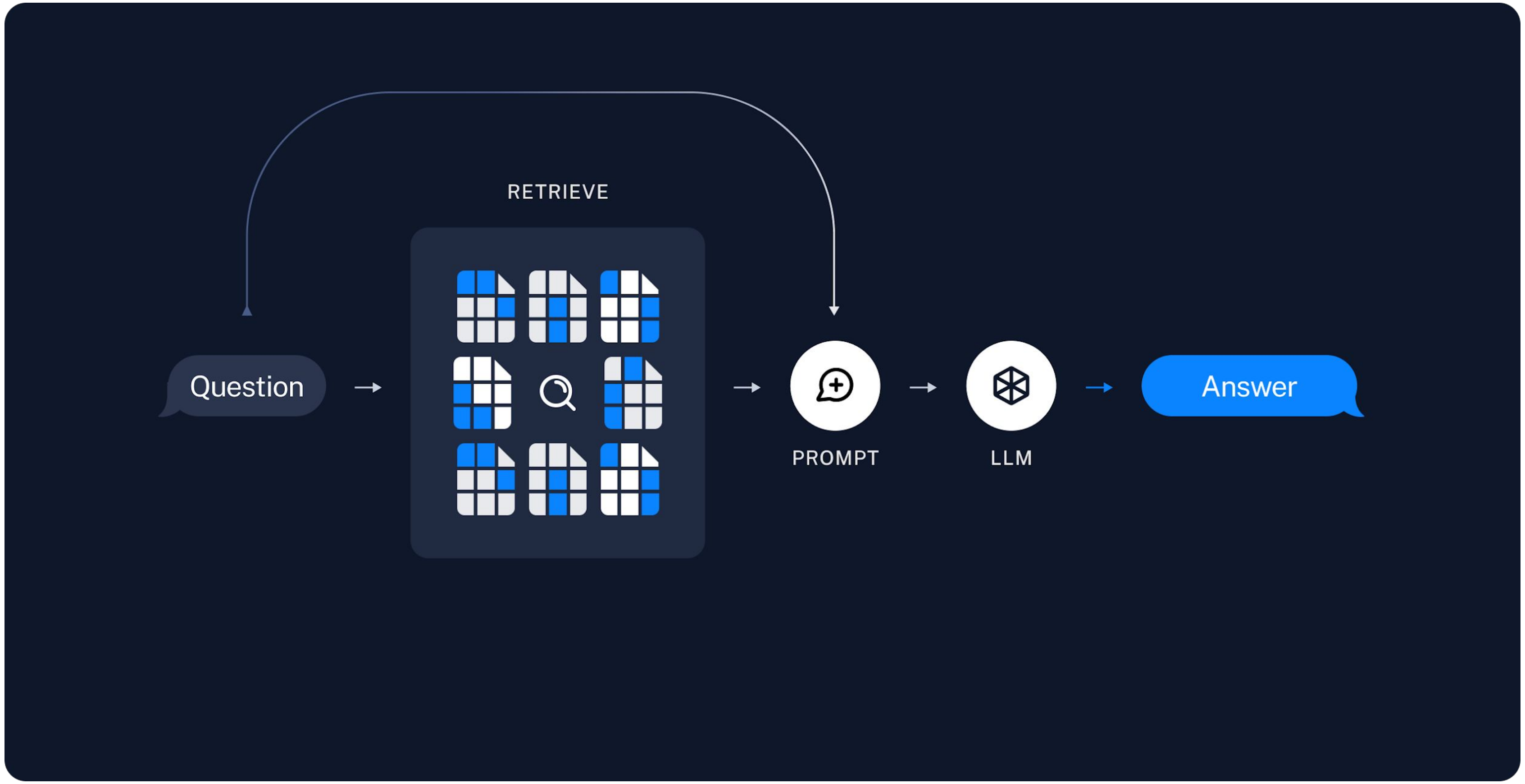
Within the Faculty of Environmental Sciences, some research groups focus on geographic information science themes, for example:

- The Chair of Cartographic Communication: core research interests include map generalization, geovisualization, and volunteered geographic information. Among others, the research group is currently actively involved in the MSc Cartography, an Erasmus Mundus Master of Science Programme that provides a broad education in Cartography and Geoinformatics. The programme awards a joint degree from four universities: Technical University of Munich, TU Wien, TU Dresden and the University of Twente.
- The Chair of Geoinformatics: core research interests include spatial data infrastructures, spatial decision support systems, and geoprocessing on the Web. Among others, the research group is currently actively contributing to the advancement of scientific (geo)data infrastructure in Germany, through speaker roles in NFDI4Earth (National Research Data Infrastructure for the Earth System Sciences) and Base4NFDI (Basic Services for Research Data Infrastructure in Germany).

2 Programme composition

The call for papers welcomed submissions across five categories, each with its own submission process: workshops/tutorials, full papers, short papers, posters, and published articles. The number of submissions in each category was: 9 workshop/tutorial proposals, 30 full papers,

RAG: Retrieval and Generation

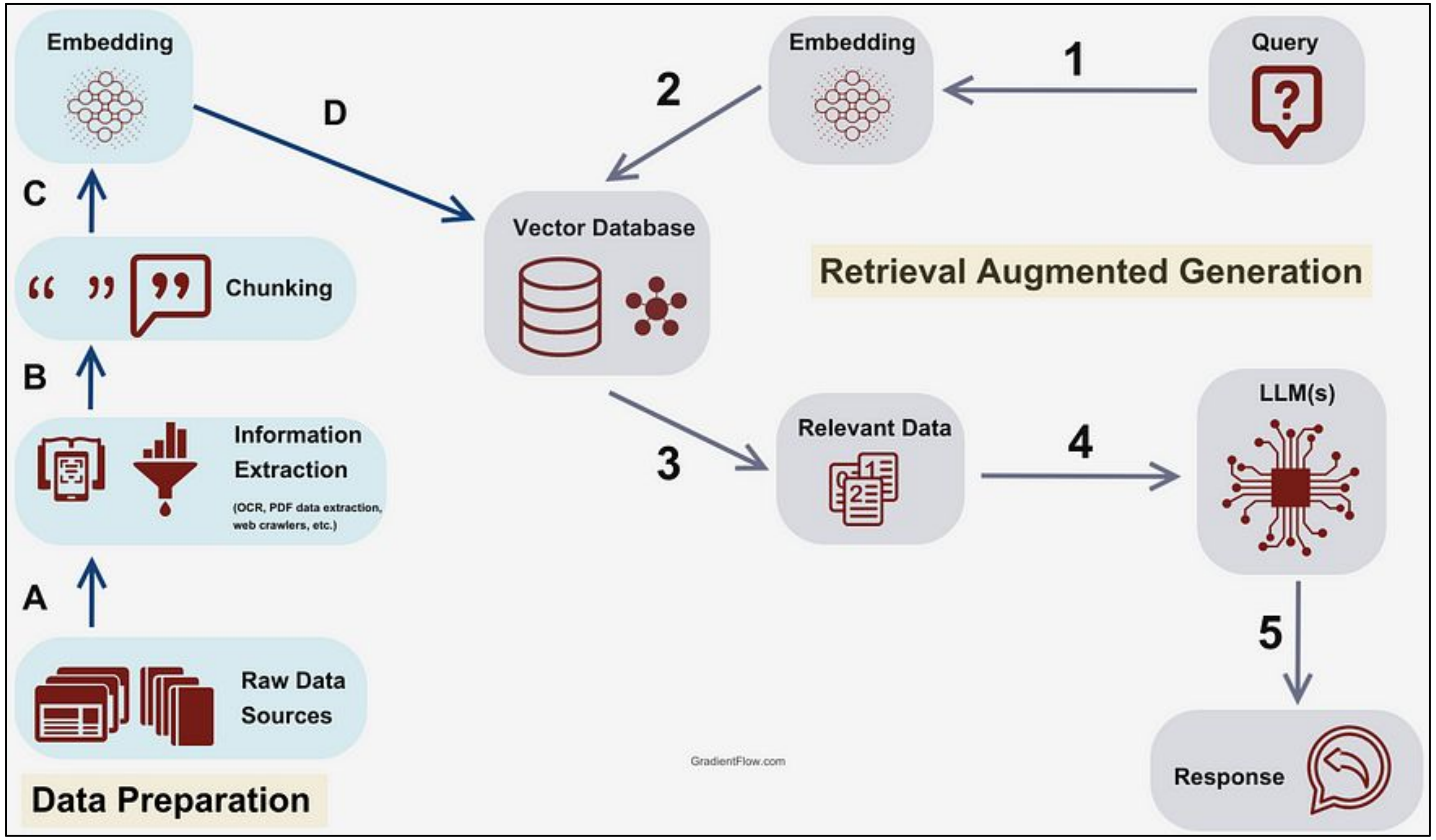


https://python.langchain.com/docs/tutorials/rag/

Q: “Who are the authors of this document?”

A: “The authors are…”

RAG: All steps at once

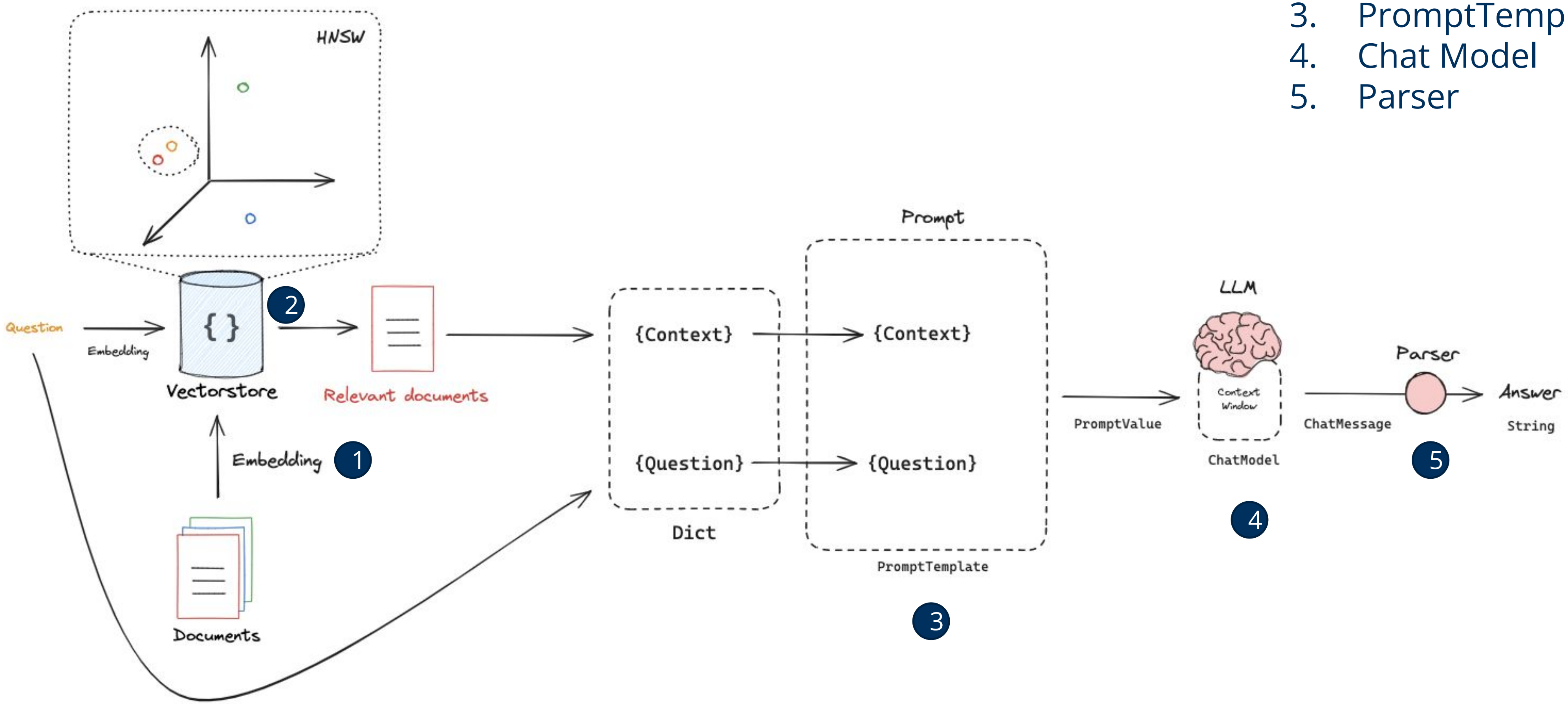


https://aswin19031997.medium.com/enhancing-conversational-ai-with-retrieval-augmented-generation-rag-leveraging-csv-integration-3000322819eb

Examples of RAG Architectures

Basic RAG Architecture

1. Embedding Model
2. Vector Store
3. PromptTemplate
4. Chat Model
5. Parser



Our scenario: from search target to embedding

all metadata


```
{ 'tourism': 'museum',
 'building': 'yes',
 'addr:postcode': '01069',
 'blind:description:de': 'Führunge für Sehgeschädigte und Blinde',
 'fee': 'yes',
 'email': 'service@dhmd.de',
 'addr:housenumber': '1',
 'id': "('relation', 5651)",
 'addr:street': 'Lingnerplatz',
 'heritage': 'yes',
 'dog': 'no',
 'check_date': '2024-03-09',
 'website': 'https://www.dhmd.de',
 'internet_access': 'wlan',
 'name:en': 'German Hygiene Museum',
 'wikidata': 'Q874373',
 'roof:shape': 'flat',
 'roof:colour': '#727466',
 'alt_name:de': 'Hygienemuseum',
 'toilets': 'yes',
 'toilets:wheelchair': 'yes',
 'internet_access:fee': 'no',
 'building:architecture': 'neoclassicism',
 'type': 'multipolygon',
 'wikipedia': 'de:Deutsches Hygiene-Museum',
 'architect': 'Wilhelm Kreis',
 'changing_table': 'yes',
 'building_description': 'yes',
 'wheelchair': 'yes',
 'opening_hours': 'Tu-Su 10:00-18:00',
 'name': 'Deutsches Hygiene-Museum',
 'geometry': '{"type": "Polygon", "coordinates": [[[13.7455854, 51.
 'internet_access:access': 'customers',
 'check_date:opening_hours': '2022-06-20',
 'start_date': '1928..1930',
 'smoking': 'no',
 'addr:city': 'Dresden',
 'phone': '+49 351 4846400'}
```


selected metadata

Text-based metadata only

Additional examples

```
Document 1:
Name: Bürgerhaus Schönborn
addr:city: Dresden
addr:country: DE
addr:housenumber: 6
addr:postcode: 01465
addr:street: Seifersdorfer Straße
amenity: townhall
building: yes
source: HiRes aerial imagery
building_description: yes

--------------------
Document 2:
Name: A-Gebäude
building: yes
building_description: yes

--------------------
Document 3:
Name: B-Gebäude
building: yes
building_description: yes

--------------------
Document 4:
Name: Geschwisterwohnen WG
building: yes
operator: Outlaw
website: https://www.outlaw-ggmbh.de/wo
building_description: yes
```

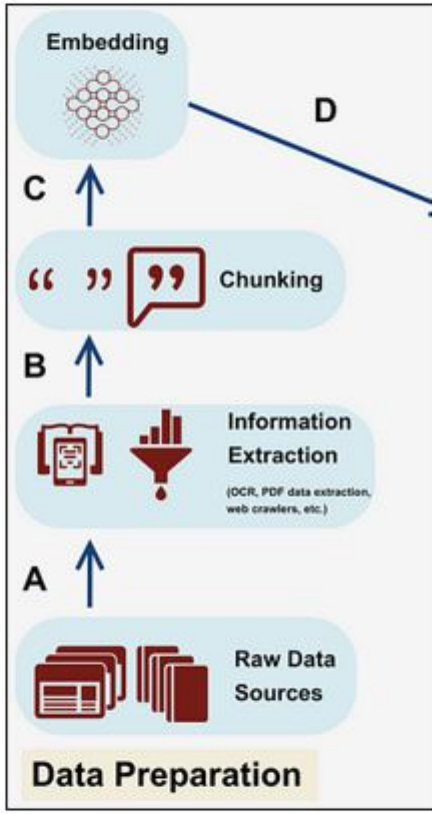
embedding

“An embedding is a numerical representation of a piece of information, for example, text, documents, images, audio” [source]

```
[[-0.02388945  0.05525852 -0.01165488 ...
 [-0.0126876   0.04687412 -0.01050217 ...
 [ 0.00049438  0.11941205  0.00522949 ...
 ...
 [-0.03900796 -0.01060951 -0.00738271 ...
 [-0.09598278 -0.06301168 -0.11690582 ...
 [-0.01162949  0.05961934  0.01650903 ...
```

DRESDEN concept

Embedding model for vector creation



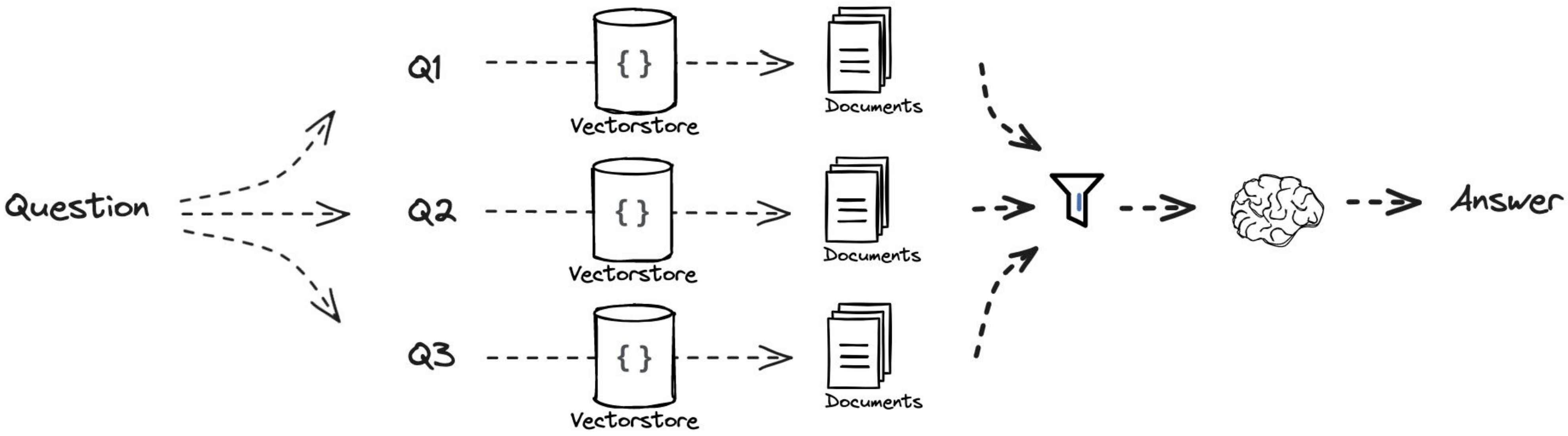

```
from chromadb.utils.embedding_functions  import SentenceTransformerEmbeddingFunction

model_name = "paraphrase-multilingual-mpnet-base-v2"

ef = SentenceTransformerEmbeddingFunction(model_name=model_name)
```

Examples of RAG Architectures

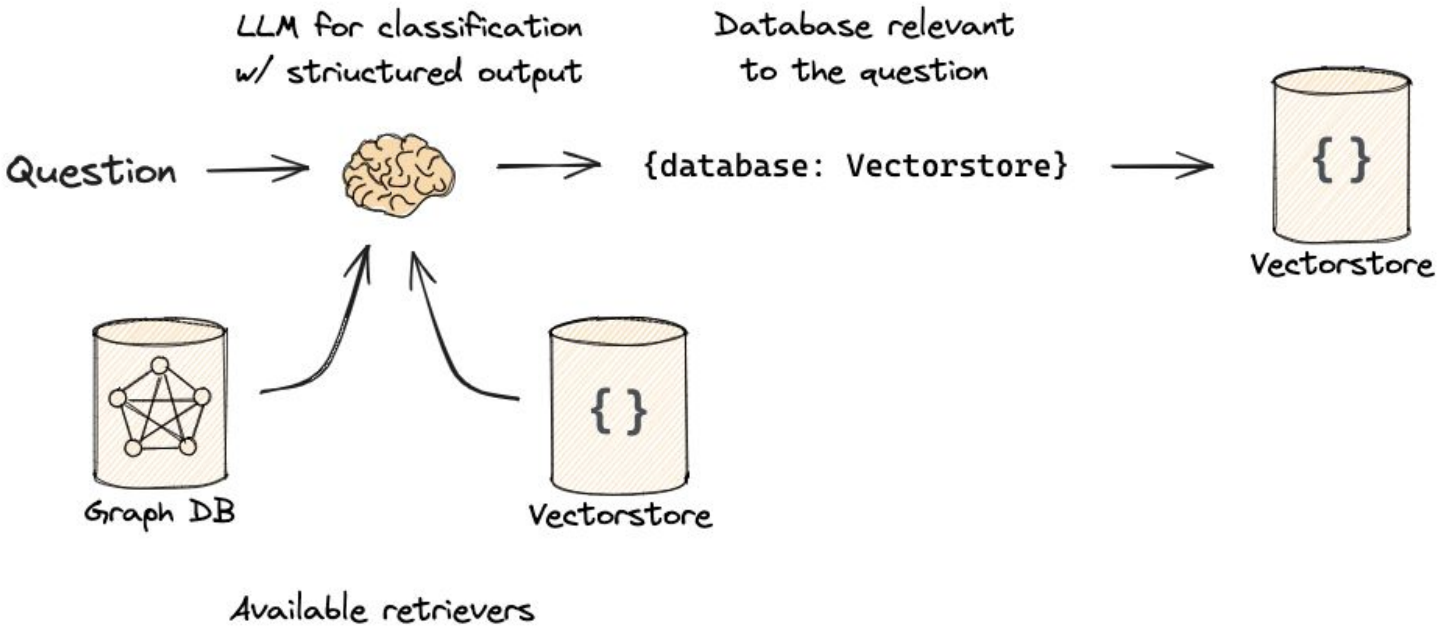
Expansion: RAG Fusion



https://github.com/langchain-ai/rag-from-scratch/blob/main/rag_from_scratch_5_to_9.ipynb

Examples of RAG Architectures

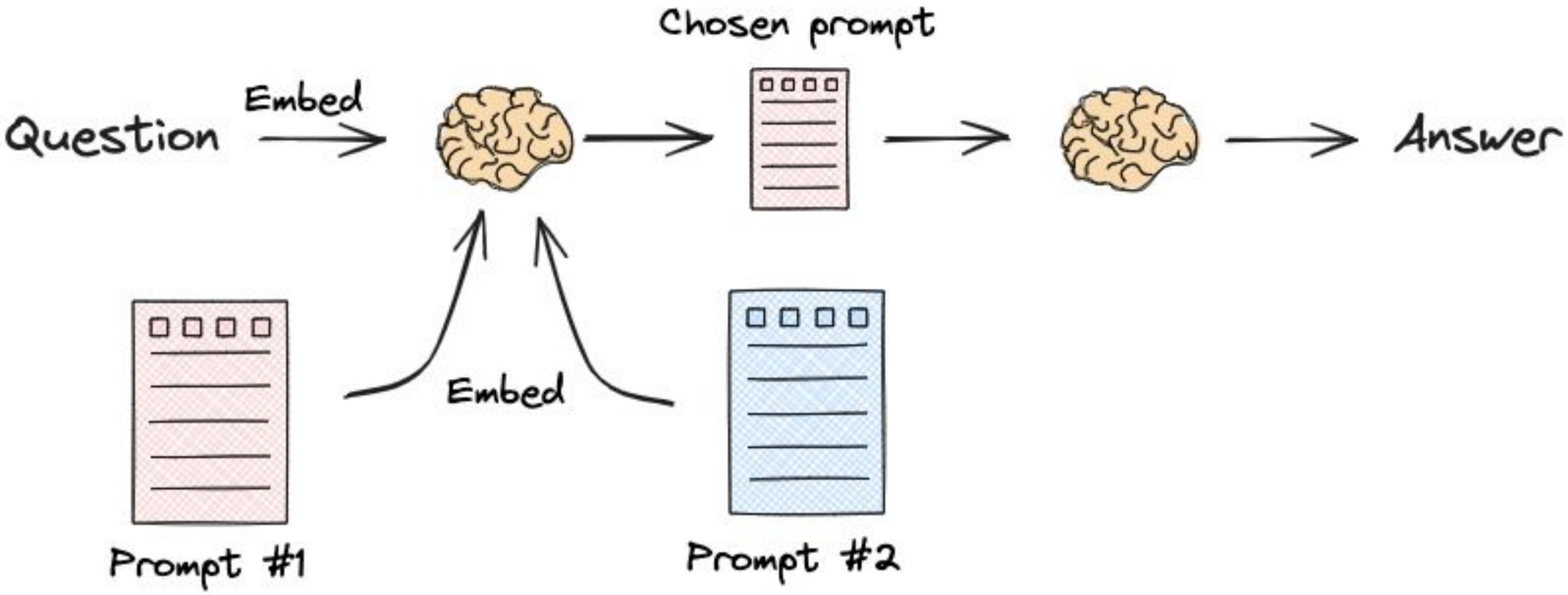
Logical Routing: Let LLM choose DB based on the question



https://github.com/langchain-ai/rag-from-scratch/blob/main/rag_from_scratch_10_and_11.ipynb

Examples of RAG Architectures

Semantic Routing: Embed question and choose prompt based on similarity



https://github.com/langchain-ai/rag-from-scratch/blob/main/rag_from_scratch_10_and_11.ipynb

References

- Baeza-Yates, R. and Ribeiro-Neto, B. (1999) *Modern information retrieval*. New York, New York, USA: ACM Press.
- Carniel, A.C. (2023) 'Defining and designing spatial queries: the role of spatial relationships', *Geo-spatial Information Science*, pp. 1–25. Available at: https://doi.org/10.1080/10095020.2022.2163924.
- Degbelo, A. (2022) 'FAIR geovisualizations: definitions, challenges, and the road ahead', *International Journal of Geographical Information Science*, 36(6), pp. 1059–1099. Available at: https://doi.org/10.1080/13658816.2021.1983579.
- Dimitrakis, E., Sgontzos, K. and Tzitzikas, Y. (2020) 'A survey on question answering systems over linked data and documents', *Journal of Intelligent Information Systems*, 55(2), pp. 233–259. Available at: https://doi.org/10.1007/s10844-019-00584-7.
- Fu, G., Jones, C.B. and Abdelmoty, A.I. (2005) 'Ontology-based spatial query expansion in information retrieval', in R. Meersman, Z. Tari, M.-S. Hacid, J. Mylopoulos, B. Pernici, Ö. Babaoglu, H.-A. Jacobsen, J.P. Loyall, M. Kifer, and S. Spaccapietra (eds) *On the Move to Meaningful Internet Systems 2005*. Springer, pp. 1466–1482. Available at: https://doi.org/10.1007/11575801_33.
- Purves, R.S., Clough, P., Jones, C.B., Hall, M.H. and Murdock, V. (2018) 'Geographic information retrieval: progress and challenges in spatial search of text', Foundations and Trends® in Information Retrieval, 12(2–3), pp. 164–318. Available at: https://doi.org/10.1561/1500000034.
- Zaib, M., Zhang, W.E., Sheng, Q.Z., Mahmood, A. and Zhang, Y. (2022) 'Conversational question answering: a survey', *Knowledge and Information Systems*, 64(12), pp. 3151–3195. Available at: https://doi.org/10.1007/s10115-022-01744-y.