UNICAM

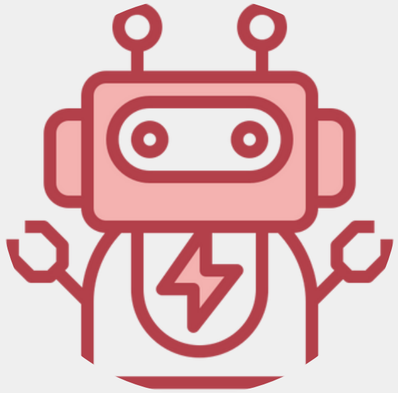
Technology for Big Data

Memory Store

Instructor: Prof.Massimo Callisto De Donato
Student: Van Hieu Ho

Outline

Introduction



Technology and Function

Further Development

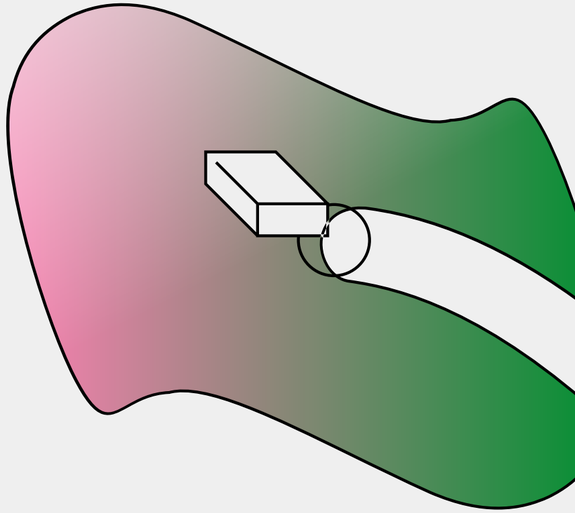
Testing

Introduction

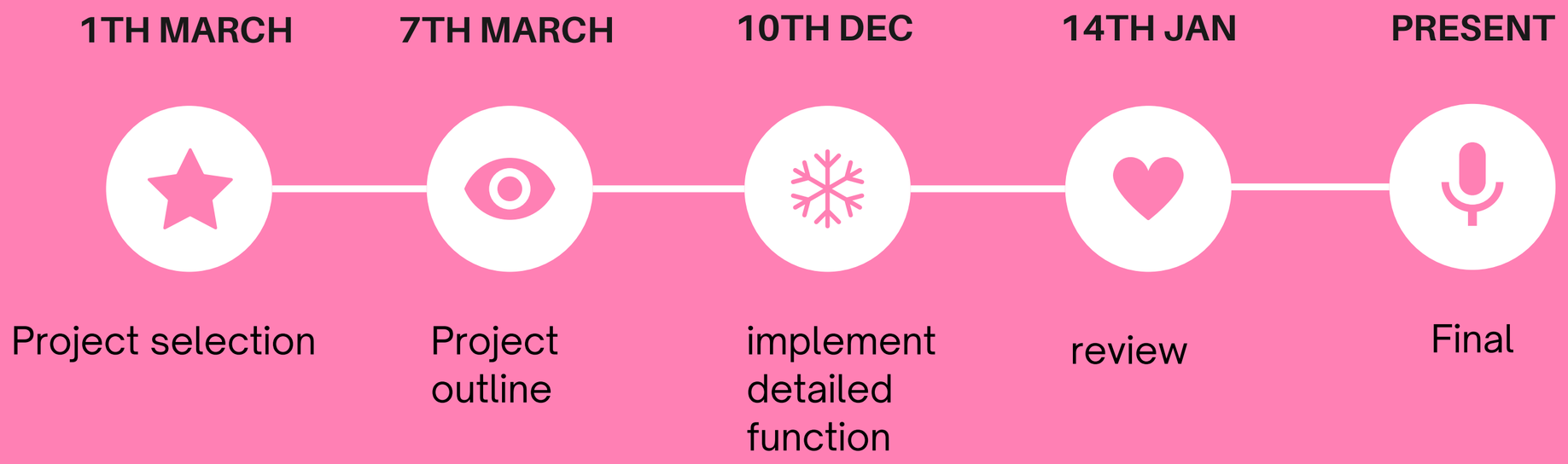
Overview of the project and its purpose

Main objectives:

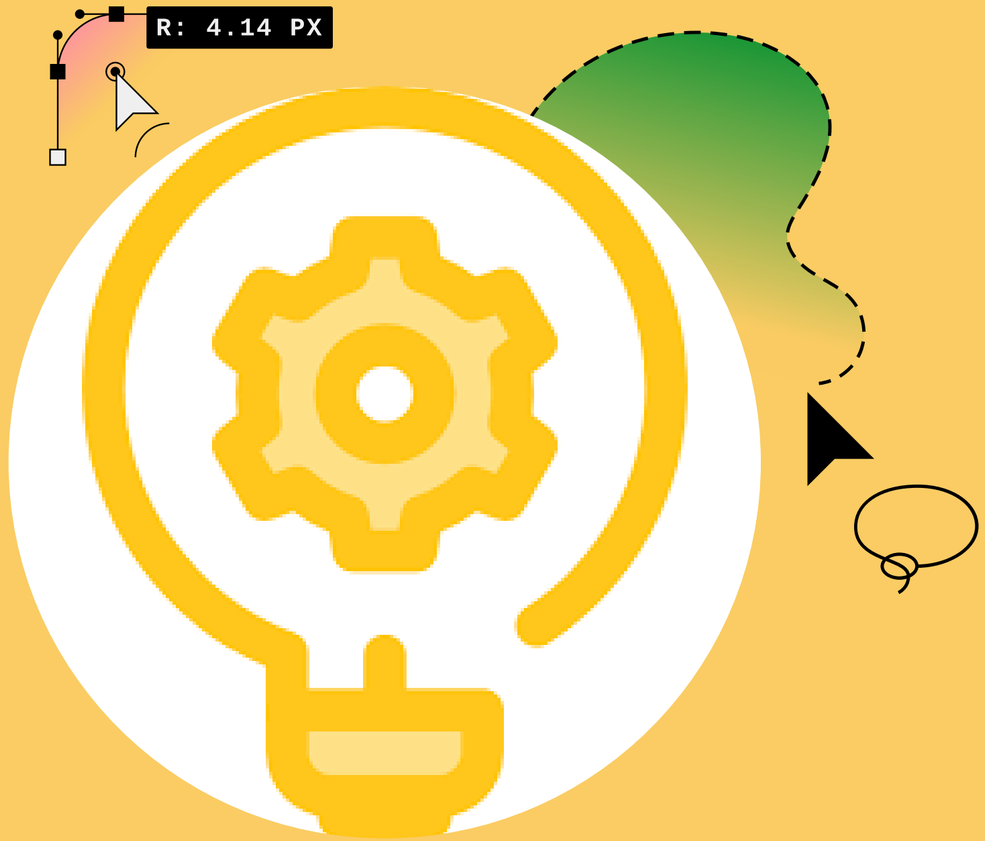
- Collect data from CEUR.
- Store it in MongoDB.
- Visualize the data.
- Implement advanced search using LLM-Vector Search.



Development Phase



Technology and Function



2.Technology

2.Function, Explaination

2.Technology Used

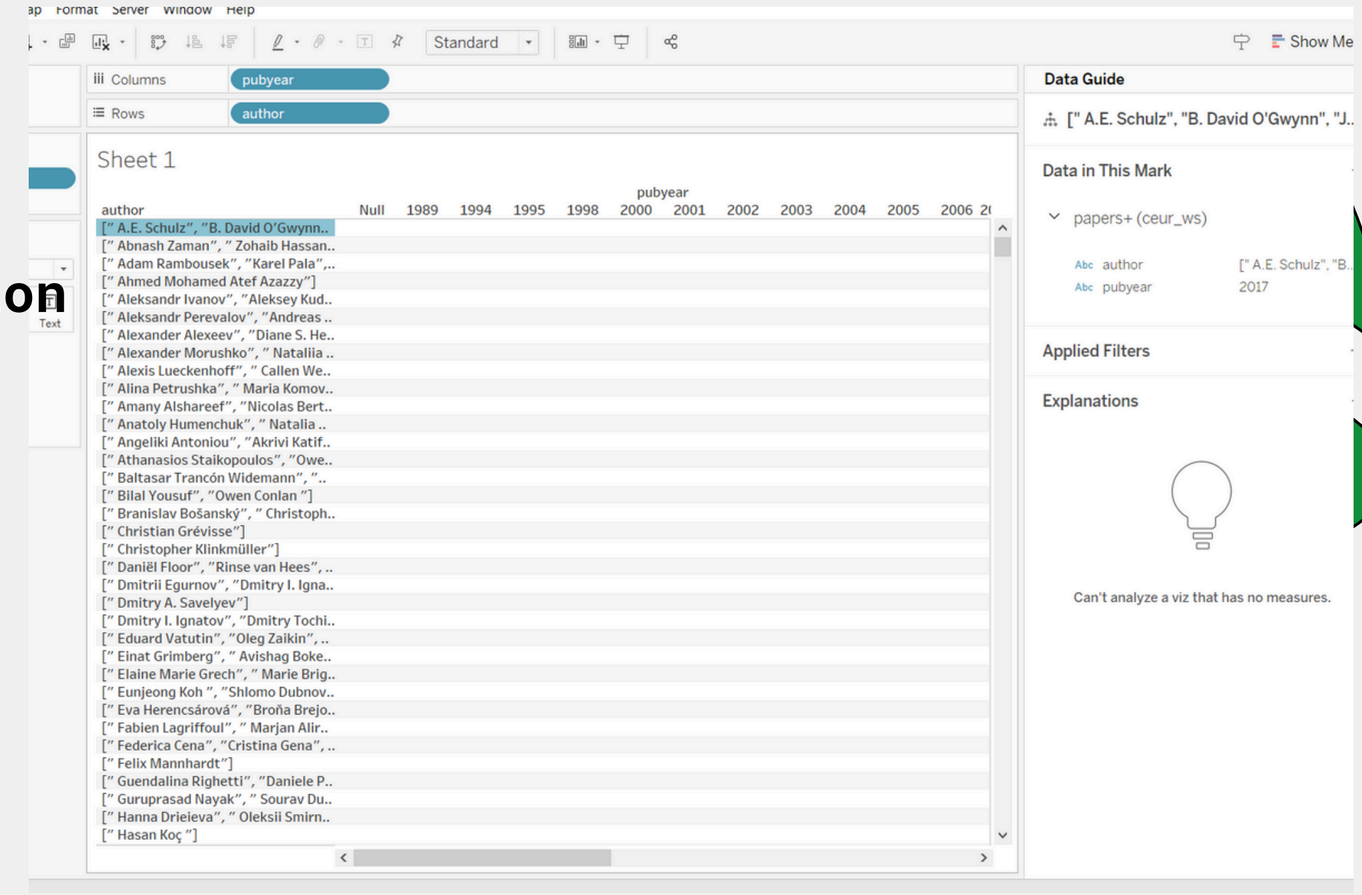
1.Data Collection & Storage

- Crawl data from CEUR and store it in MongoDB.

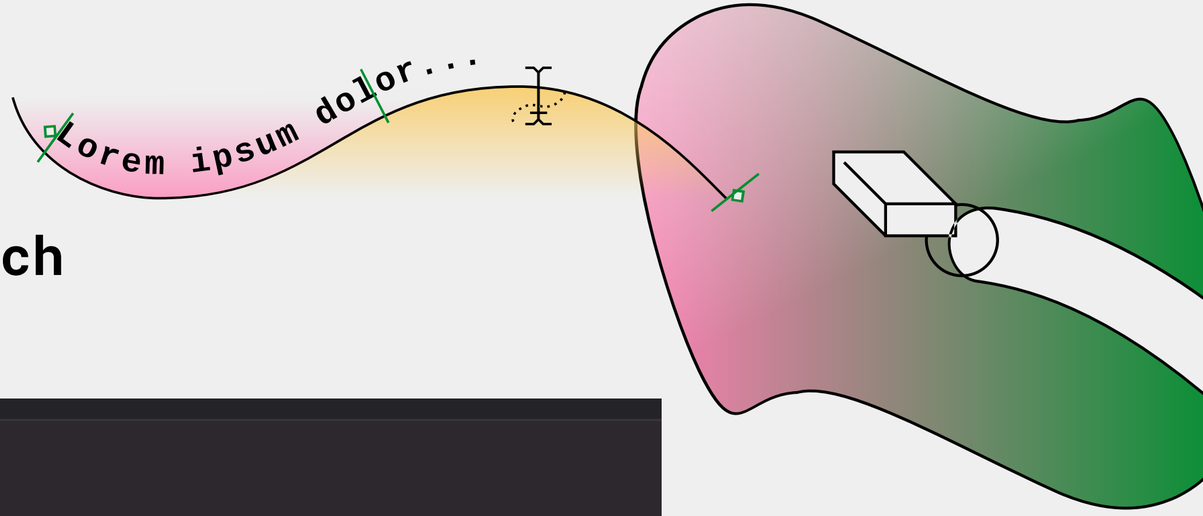
2.Technology Used

1.Data Processing & Transformation

Convert and preprocess data for visualization.



2.Technology Used



1.Query Data based on llm-vector-search

```
Pinged your deployment. You successfully connected to MongoDB!
```

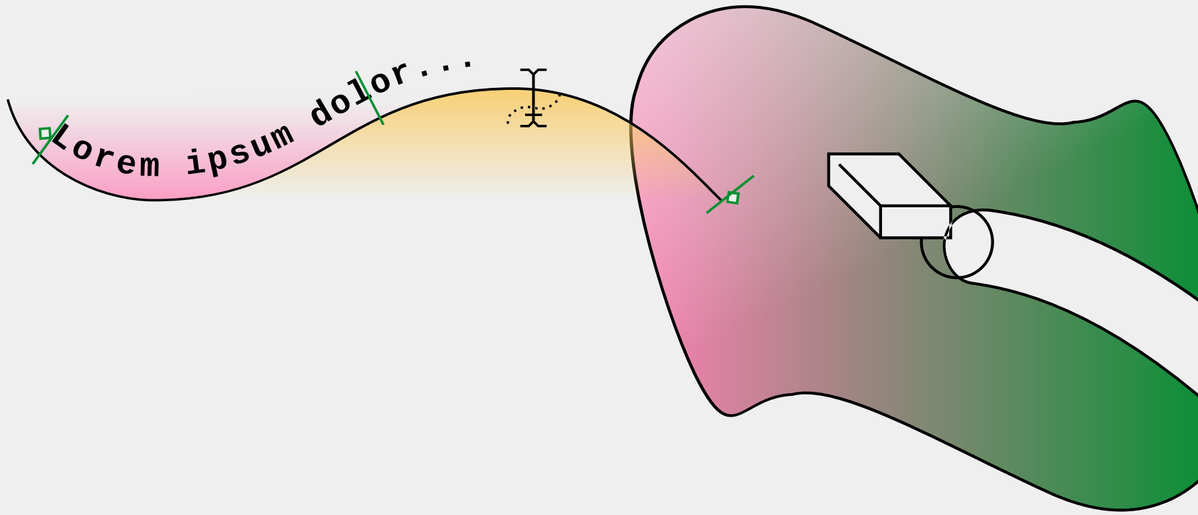
```
# Fetch the first document from the collection
db = client.get_database('ceur_ws')
papers_collection = db.papers

title = "Empowering Supply Chains Resilience: LLMs-Powered BN for Proactive Supply Chain Risk Identification"
document = papers_collection.find_one({'title': title})
print(document)
```

```
{'_id': ObjectId('67d7dec76dba08ca73f92681'), 'url': 'https://ceur-ws.org/Vol-3707/D2R224_paper_2.pdf', 'title': 'Empowering Supply
```

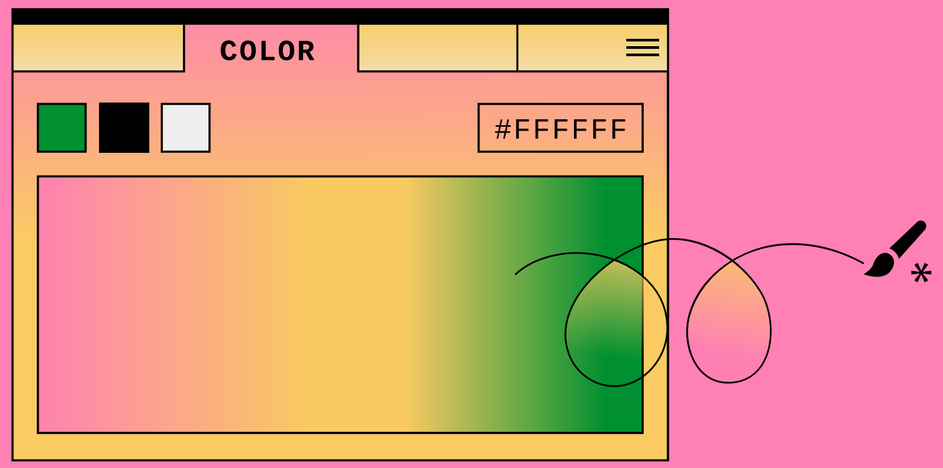
2.Technology Used

1.Data Processing & Transformation



- Implementing vector-based retrieval for semantic search.
- How LLM and embeddings improve search accuracy.

Further Development



1.Fully system with orchestration function

2.Advanced Search with LLM-Vector Search:

- Implementing intelligent search with Vector Search.

Testing and Question