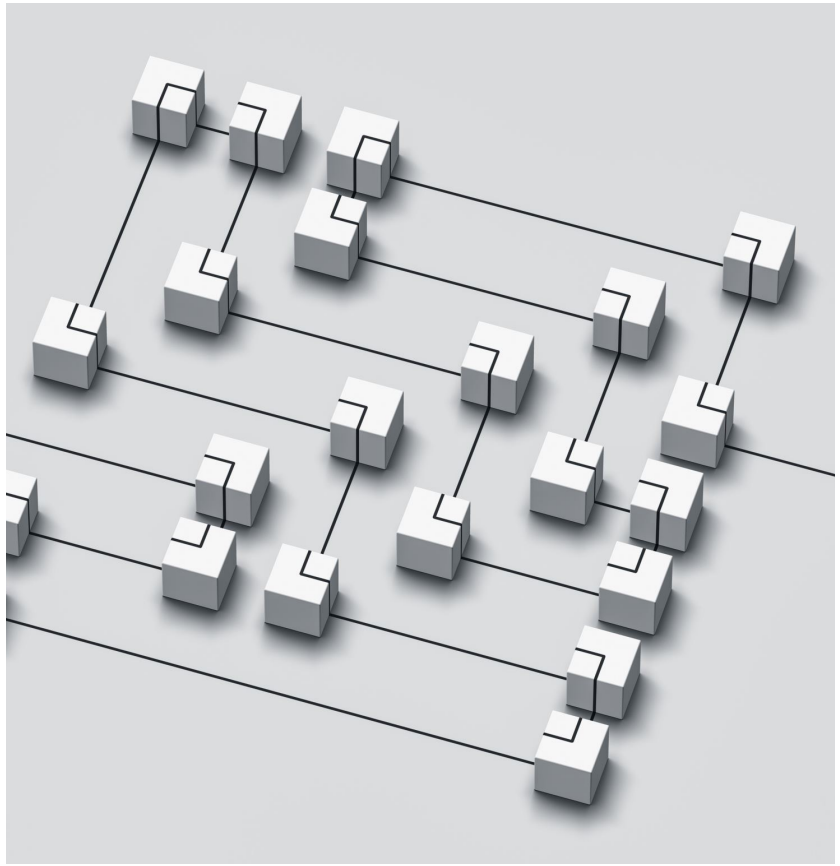
Keyword and Semantic Search with Rerank

Maryam Zubair

Table of Content

Introduction

This project leverages advanced techniques in the field of natural language processing (NLP) and semantic search to significantly enhance the quality of search results. It does so by integrating two powerful APIs: Weaviate and Cohere.

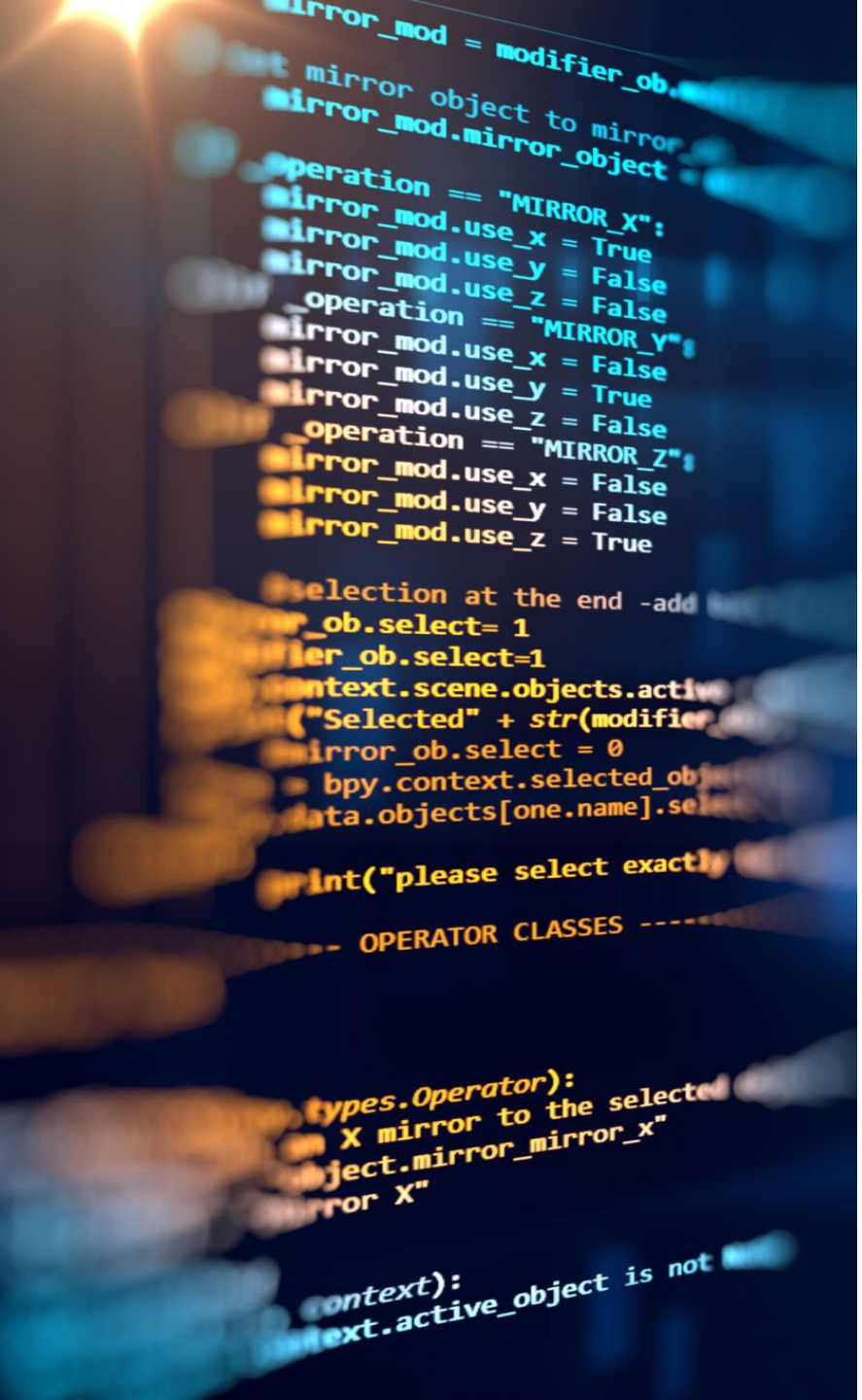


API

1. **Weaviate:**
    1. Weaviate is an open-source search engine powered by machine learning, specifically designed for scalable and vectorized data retrieval.
    2. In the context of this project, Weaviate is used to facilitate semantic search and keyword-based retrieval, providing a more sophisticated method of sifting through data based on the meaning and context of the search queries.
2. **Cohere :**
    1. Cohere provides powerful natural language understanding capabilities. It's used in this project to rerank search results, which means adjusting the order of search results so that the most relevant ones appear first.
    2. This reranking is based on the comprehension of the text, making it more aligned with the user’s search intent.

Semantic Search Techniques:

1. Unlike traditional keyword-based search, semantic search seeks to understand the searcher's intent and the contextual meaning of terms as they appear in the searchable dataspace.
2. This approach allows for more accurate and relevant search results, as the system can interpret the query in a more human-like manner.



Combination of Keyword Search, Dense Retrieval, and Reranking Mechanisms:

1. **Keyword Search:** Traditional method of finding documents or data that contain specific words or phrases.
2. **Dense Retrieval:** A modern approach in NLP that retrieves information based on semantic representations. Instead of matching exact words, it understands the meaning behind the words.
3. **Reranking Mechanisms:** After the initial retrieval of data, reranking reorders these results to prioritize the ones most likely to be relevant to the query.

Design

1. **Cohere API:**
    - Implementing the Cohere API.
    - Setting Up the Cohere API Key.
    - Initializing the Cohere API.

```python
import os
import cohere
# Set the API key as an environment variable
os.environ['COHERE_API_KEY'] = 'cohere-key'

# Now you can use it to instantiate the Cohere client
co = cohere.Client(api_key=os.environ['COHERE_API_KEY'])
```

Design(cont..)

1. **Weaviate API:**
    - Setting Up the Weaviate API Key.
    - Utilizing the Weaviate API.
    - Initializing the Weaviate API.

```
import os
import weaviate

# Set the API key as an environment variable
os.environ['WEAVIATE_API_KEY'] = 'weaviate-default-key' # OR default key 7

# Now you can use it to instantiate the auth configuration
auth_config = weaviate.auth.AuthApiKey(api_key=os.environ['WEAVIATE_API_KE
```

Implementation

1. Setup API: Configuration of API keys and installation of necessary packages.
2. Weaviate Integration: Description of keyword search and dense retrieval functions.
3. Cohere Integration: Description of reranking search results function.

Implementation(cont..)

4. Keyword Search and Dense Retrieval: Implementation details and output.

```
[11] dense_retrieval_results = dense_retrieval(query, client)

     print_result(dense_retrieval_results)

views:2000


item 1
_additional:{'distance': -146.58931}

lang:en

text:China had one of the largest economies in the world for most of the past two thousand years, during which it has seen cycles of prosperity and decline.

title:China

url:https://en.wikipedia.org/wiki?curid=5405

views:4000


item 2
_additional:{'distance': -146.22166}

lang:en

text:The Liao fell to the Jurchen Jin dynasty in 1122, which gave the city to the Song dynasty and then retook it in 1125 during its conquest of northern Chi

title:Beijing

url:https://en.wikipedia.org/wiki?curid=18603746
```

Implementation (cont..)

5. Reranking: Process and analysis of reranked results.

```
[24] results = dense_retrieval(query_2,client)

for i, result in enumerate(results):
    print(f"i:{i}")
    print(result.get('title'))
    print(result.get('text'))
    print()
```

```
i:0
Robert Wadlow
Robert Pershing Wadlow (February 22, 1918 July 15, 1940), also known as the Alton Giant and the Giant of Illinois, was a man who was the tallest person in r

i:1
Manute Bol
Bol came from a family of extraordinarily tall men and women. He said: "My mother was , my father , and my sister is . And my great-grandfather was even tal

i:2
Sultan Kösen
Sultan Kösen (born 10 December 1982) is a Turkish farmer who holds the Guinness World Record for tallest living male at . Of Kurdish ethnicity, he is the se

i:3
Sultan Kösen
Kösen turned 40 years old on 10 December 2022. He celebrated his birthday a few days early by visiting the Ripley's Believe It or Not! museum in Orlando, Fl

i:4
Netherlands
The Dutch are the tallest people in the world, by nationality, with an average height of for adult males and for adult females in 2009. The average height o
```

Testing

```
for i, rerank_result in enumerate(reranked_text):
    print(f"i:{i}")
    print(f"{rerank_result}")
    print()
```

```
i:0
RerankResult<document['text']: Robert Pershing Wadlow (February 22, 1918 July 15, 1940), also known as the Alton Giant and the Giant of Illinois, was a man v

i:1
RerankResult<document['text']: Sultan Kösen (born 10 December 1982) is a Turkish farmer who holds the Guinness World Record for tallest living male at . Of h

i:2
RerankResult<document['text']: The Dutch are the tallest people in the world, by nationality, with an average height of for adult males and for adult females

i:3
RerankResult<document['text']: Kösen turned 40 years old on 10 December 2022. He celebrated his birthday a few days early by visiting the Ripley's Believe It

i:4
RerankResult<document['text']: Bol came from a family of extraordinarily tall men and women. He said: "My mother was , my father , and my sister is . And my
```

Conclusion

This project synergizes Cohere and Weaviate APIs, transcending traditional search limitations. By combining advanced NLP and semantic search, it delivers nuanced, context-aware results. The modular design ensures easy integration, serving as a blueprint for superior search application

GitHub Link

https://github.com/Maryam-Zubair/MachineLearning_Assignment/tree/main/ChatGPT/Rerank