

# Keyword and Semantic Search with ReRank

Using Fine Tune, OpenAI,  
Semantic search

# Table of content

1. Introduction
2. Design
3. Implementation
4. Test
5. Conclusion
6. Github repository Link:
7. Bibliography / References

# Introduction

This project utilizes natural language processing and semantic search techniques to enhance search results using Weaviate and Cohere APIs. The primary goal is to improve the relevance of search queries through a combination of keyword search, dense retrieval, and reranking mechanisms.

# Design: Technology Used

## 1. Cohere API:

- Utilized for reranking search results.
- API key is set up as an environment variable.
- Cohere client instantiated using the Python ``cohere`` package.

## 2. Weaviate API:

- Employed for semantic search and keyword-based retrieval.
- API key and configuration set up as environment variables.
- Weaviate client instantiated using the Python ``weaviate`` package.

## 3. Python Packages:

- ``python-dotenv`` for managing environment variables.
- Custom utility functions stored in a ``utils.py`` file for modular code organization.

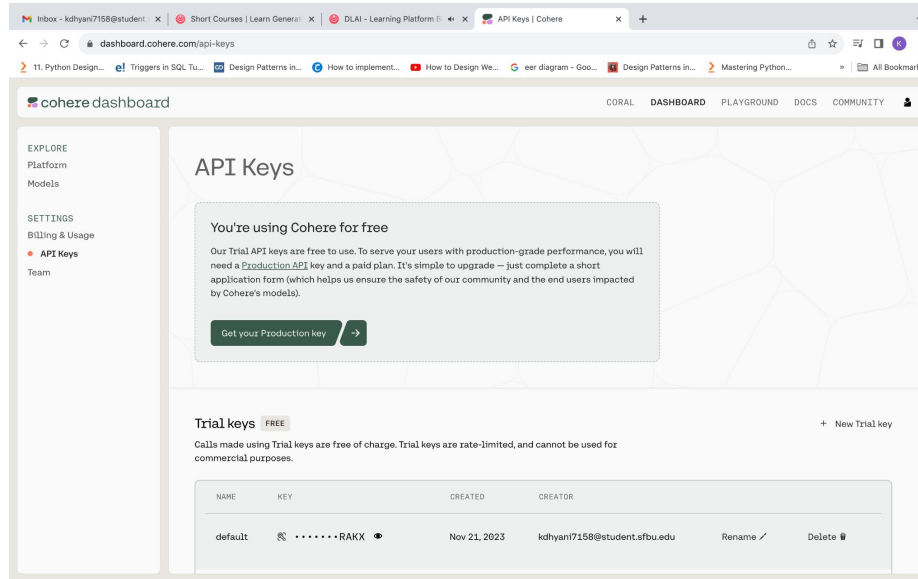
# Implementation : Used Google colab

## 1. Setup API

- API keys for Cohere, Weaviate, and potentially OpenAI are configured as environment variables.
- Necessary packages (`cohere`, `weaviate-client`, `python-dotenv`) are installed.

### *Cohere API key information:*

Website: <https://dashboard.cohere.com/> create a account and they give you trial API key. copy the key.



# cont..

- **Weaviate API key:** you can create account and get the API key. However I used universal API key '76320a90-53d8-42bc-b41d-678647c6672e' & Universal URL: '<https://cohere-demo.weaviate.network/>'.

**NOTE: IF your Above two mentioned key is not set up then its gives you error**

- **Install the following packages**

!pip install cohere

!pip install weaviate-client

! pip install python-dotenv

## 2. Weaviate Integration:

- Weaviate is used for both keyword search and dense retrieval.
- A utility function `dense\_retrieval` is defined to perform dense retrieval given a query.

## 3. Cohere Integration:

- The Cohere API is employed for reranking search results.
- A utility function `rerank\_responses` is defined to rerank a list of responses using the Cohere API

# cont.

## 4. Keyword Search:

- The `keyword\_search` function is implemented to perform keyword-based searches using Weaviate.
- Results are printed with customizable properties and additional information.

## 5. Dense Retrieval:

- The `dense\_retrieval` function is implemented to perform dense retrieval using Weaviate.
- Results are printed with customizable properties and additional information.

```
[22] query = "What is the capital of Canada?"

[23] dense_retrieval_results = dense_retrieval(query, client)

print_result(dense_retrieval_results)

item 0
_additional: {'distance': -150.8031}
lang: en
text: The governor general of the province had designated Kingston as the capital in 1841. However, the major population centres of Toronto and Montreal, as w
title: Ottawa
url: https://en.wikipedia.org/wiki?curid=22219
views: 2000

item 1
_additional: {'distance': -150.28354}
lang: en
text: For brief periods, Toronto was twice the capital of the united Province of Canada: first from 1849 to 1852, following unrest in Montreal, and later 1856
title: Toronto
url: https://en.wikipedia.org/wiki?curid=64646
views: 3000

item 2
```

# cont..

## 6. Reranking:

- Responses obtained from keyword search or dense retrieval are reranked using the Cohere API.
- Reranked results are printed for analysis.

### Improving Keyword Search with ReRank

```
[25] query_1 = "What is the capital of Canada?"
```

```
query_1 = "What is the capital of Canada?"
results = keyword_search(query_1,
                          client,
                          properties=["text", "title", "url", "views", "lang", "_additional {distance}"],
                          num_results=3)

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

```
i:0
Monarchy of Canada
In his 1990 book, "Continental Divide: the Values and Institutions of the United States and Canada," Seymour Martin Lipset argues that the presence of the mo
i:1
Early modern period
North America outside the zone of Spanish settlement was a contested area in the 17th century. Spain had founded small settlements in Florida and Georgia but
i:2
Flag of Canada
By the Second World War, the Red Ensign was viewed as Canada's "de facto" national flag. A joint committee of the Senate and House of Commons was appointed o
```

[uses keyword search: gives you more than 500 item]

### Improving Dense Retrieval with ReRank

```
[31] query_2 = "Who is the tallest person in history?"
```

```
[32] 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 re
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 tall
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 sev
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, Flo
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 of
```

[uses Dense Retrieval : gives you few relevant response]



# Test

```
texts = [result.get('text') for result in results]
reranked_text = rerank_responses(query_2, texts)
```

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

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 K

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 applications.

# Github Link

<https://github.com/DKruti/Machine-Learning/tree/master/Generative%20AI/FINE%20TUNING/%20keyword%20and%20semantic%20search%20with%20ReRank>

# References

- [https://hc.labnet.sfbu.edu/~henry/sfbu/course/deeplearning\\_ai/llm\\_semantic\\_search/slides/exercise\\_llm\\_semantic\\_search.html](https://hc.labnet.sfbu.edu/~henry/sfbu/course/deeplearning_ai/llm_semantic_search/slides/exercise_llm_semantic_search.html)
- <https://learn.deeplearning.ai/large-language-models-semantic-search/lesson/5/rerank>