

# Keyword Search and Dense Retrieval with Rerank

Gangzhaorige Li

github

# Project Overview Process

1. Import cohere and weaviate
2. Apply Dense retrieval to a query.
3. Improve Keyword search with Rerank.
4. Improve Dense Retrieval with Rerank,

# Dense Retrieval

Dense vectors:

- In dense retrieval, documents and queries are represented as dense vectors in a high-dimensional vector space.
- Each dimension of the vector carries information about different aspects of the text.

Semantic information:

- Dense vectors capture rich semantic information about the text, allowing for more nuanced understanding of document-query relevance.

Embeddings:

- Dense retrieval often involves using word embeddings or document embeddings, such as Word2Vec, Doc2Vec, or more advanced models like BERT, to represent text as dense vectors.

# Keyword Search and Rerank

1. Keyword Search: A traditional search method that relies on keywords or terms to retrieve matching documents or data.
2. Rerank: A process used to improve the performance of both Dense Retrieval and Keyword Search by reordering or re-ranking search results to enhance relevance and accuracy.

# Keyword Search and Rerank

1. Keyword Search: A traditional search method that relies on keywords or terms to retrieve matching documents or data.
2. Rerank: A process used to improve the performance of both Dense Retrieval and Keyword Search by reordering or re-ranking search results to enhance relevance and accuracy.

# Requirements

Requirements: cohere and weaviate

Cohere: A platform or system designed for various natural language processing (NLP) and text-related tasks.

Weaviate: An open-source vector search engine used for storing, searching, and retrieving data vectors.

# Result(Apply Dense retrieval):

```
query = "What is the capital of Canada?"  
dense_retrieval_results = dense_retrieval(query, client)  
print_result(dense_retrieval_results)
```

```
item 0  
  additional: {'distance': -150.82233}
```

```
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 well as the former capital of Lower Canada, Quebec City, all had legislators dissatisfied with Kingston.
```

# Result(improving keyword search with rerank):

```
query_1 = "What is the capital of Canada?"
results = keyword_search(query_1,
                          client,
                          properties=["text", "title", "url",
                                     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 monarchy in Canada helps distinguish Canadian identity from American identity. Since at least the 1930s, supporters of the Crown have held the opinion that the Canadian monarch is also one of the rare unified elements of Canadian society, focusing both "the historic consciousness of the nation" and various forms of patriotism and national love "[on] the point around which coheres the nation's sense of a continuing personality". Former Governor General Vincent Massey articulated in 1967 that the monarchy "is part of ourselves. It is linked in a very special way with our national life. It stands for qualities and institutions which mean Canada to every one of us and which for all our differences and all our variety have kept Canada Canadian." But, according to Arthur Bousfield and Gary Toffoli



# Rerank Of Keyword Search

```
def rerank_responses(query, responses, num_responses=10):  
    reranked_responses = co.rerank(  
        model = 'rerank-english-v2.0',  
        query = query,  
        documents = responses,  
        top_n = num_responses,  
    )  
    return reranked_responses
```

```
texts = [result.get('text') for result in results]  
reranked_text = rerank_responses(query_1, texts)
```

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

```
i:0  
Monarchy of Canada  
i:1  
Early modern period  
i:2  
Flag of Canada  
i:3  
Flag of Canada
```

# Improving Dense Retrieval Rerank

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

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()
```

```
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:{i}")
    print(f"{rerank_result}")
    print()
```

Church of England

i:0

Robert Wadlow

Robert Pershing Wadlow (February 22, 1918 July 15, 1940) was a man who was the tallest person in the world, and the Giant of Illinois, was a man who was the tallest person in the world for whom there is irrefutable evidence. He was born in Macks Creek, a small town in Illinois, a city near St. Louis, Missouri.

i:1

Manute Bol

Bol came from a family of extraordinarily tall men, including his father, my father, and my sister is . And my great-grandfather was a member of the group, the Dinka, and the Nilotic people of which the tallest populations in the world. Bol's hometown, Turalei, is in Sudan.