Generating Answers

Maryam Zubair

TABLE OF CONTENT

INTRODUCTION

- This initiative is centered on producing responses to queries within a predefined context, demonstrating the combined use of the Cohere API for generating embeddings and the Annoy library for effective nearest neighbor searches. The main objective is to develop a system adept at comprehending and addressing questions relevant to a particular field.

DESIGN

Input text : whole information as Text or you can pass PDF using Langchain

Chunking : split the data into parts

Embedding: generate the Vector Space

Search Index: Search using index

Query: Pass the query or question from that text

Search: it searches using semantic search with Dense retrieval

Question: follow up Question or new question

Answers: Result

IMPLEMENTATION

1. **Input:** Text and query
2. **Setup the packages**
3. **Index Creation with Annoy:** The Annoy library is employed to construct a search index from the embeddings, facilitating the quick retrieval of closely related items in the vector space.
4. **Article Search Function:** A function is developed to find articles matching specific queries using the search index and Cohere embeddings, providing pertinent context for answer formulation.
5. **Answer Generation Process:** A specialized function is crafted to produce responses to queries based on the context derived from the search, using Cohere for response generation.
6. **Utility Function Development:** Essential utility functions like print_result are created for result formatting and display.
7. **Non-Active Functions:** Some functions, such as search_wikipedia_subset and generate_given_context, are included in the script but remain unused in the current implementation

TEST

```
[19] results = ask_andrews_article(
         "What is the most viewed televised event?",
         num_generations=5
     )
```

```
for gen in results:
    print(gen)
    print('--')
```

```
 The Super Bowl is typically the most viewed televised event of the year in the United States. However, it is not mentioned in the excerpt from the article you provided.

Would you like help with anything else regarding this article?
--
 The Super Bowl is reportedly the most viewed televised event in the United States, with audiences commonly measuring in at around 100 million viewers per year. Given the promi

Would you like me to search for other commonly viewed televised events?
--
 The Super Bowl is typically the most viewed televised event in the United States. In 2017, the Big Game drew in 111.3 million viewers, which was a slight increase from the pri
--
 The Super Bowl is typically the most viewed televised event in the United States. Traditionally the event has upwards of 100 million viewers every year.

Would you like to know more about the Super Bowl?
--
 The Super Bowl is reportedly the most viewed televised event in the United States, with audiences commonly measuring in at over 100 million viewers across the country. Given t
--
```

CONCLUSION

- The "Generating Answers" initiative highlights a dynamic integration of the Cohere API and the Annoy library, presenting an efficient method for interpreting and answering queries in natural language. The project skillfully combines data processing, the creation of embeddings, and the development of a search index, illustrating a comprehensive method for improving semantic comprehension. This integration of technologies enables the system to effectively produce articulate responses rooted in contextual insights, showcasing a solid foundation for constructing advanced question-answering systems.

REFERENCES

- https://hc.labnet.sfbu.edu/~henry/sfbu/course/deeplearning_ai/llm_semantic_search/slide/exercise_llm_semantic_search.html
- https://learn.deeplearning.ai/large-language-models-semantic-search/lesson/6/generati ng-answers

GITHUB:

https://github.com/Maryam-Zubair/MachineLearning_Assignment/tree/main/ChatGPT/Generating%20Answer