Summary Report of the Assignment

Objective:

The main goal of this assignment was to:

1. Clone a Git repository containing text data.

2. Import the data into a MongoDB database.

3. Fetch specific data from the MongoDB collection based on certain criteria.

4. Index the fetched data in a Vector Store using the Faiss library.

5. Query the indexed data to retrieve relevant results.

Steps Undertaken:

1. Set Up MongoDB Cloud:

- Created a MongoDB Cloud account.

- Imported the necessary data into the translation collection.

2. Access Google Colab:

- Opened Google Colab for executing the Python scripts.

3. Connect to MongoDB Cloud from Colab:

- Installed pymongo in Colab.

- Used the connection string from MongoDB Cloud to connect to the database.

4. Fetch Translations:

- Queried the translation collection for documents where authorName is "Sri Shankaracharya" and language is "English".

- Combined the description fields into a single list of descriptions.

5. Index Content in Vector Store:

- Installed necessary packages such as transformers and faiss-cpu in Colab.

- Initialized a Vector Store and added the combined translations.

6. Query the Vector Store:

- Wrote a query script using the Faiss library to search the indexed content.

- Retrieved and printed the results.

Outcome:

By completing this assignment, the following objectives were achieved:

- Interacted with MongoDB to fetch specific data based on query parameters.

- Processed and combined data into a suitable format for indexing.

- Used the Faiss library to create a Vector Store and index the processed data.

- Queried the Vector Store to retrieve and print relevant results.