

Talk with Mongo

Talk with Mongo is a project that enables generating NoSQL MongoDB queries using Huggingface LLM and Langchain from natural language. It also provides an interface to connect to a database and query data using natural language.

Features

- Generate NoSQL queries using Huggingface LLM and Langchain.
- Create a cache database to retrieve data quickly if it has been asked before.
- Create a FAISS vector store to find similar queries. If the query is found in the cache, the response is generated from the cached database, saving an LLM call. Otherwise, an LLM call is made to generate the query, which is then run on the main database to get a response. The new query and data are added to the cache database and the new query to the vector store.

Requirements

- Python 3.7+
- MongoDB Atlas account (or local MongoDB setup)
- Huggingface Transformers
- Langchain
- FAISS

Installation

1. Clone the repository:

```
git clone https://github.com/yourusername/talk-with-mongo.git
cd talk-with-mongo
```

2. Create a virtual environment and activate it:

```
python -m venv env
source env/bin/activate # On Windows use `env\Scripts\activate`
```

3. Install the required packages:

```
pip install -r requirements.txt
```

Configuration

Update the `username` and `pwd` in the files with the MongoDB username and password and `YOUR_TOKEN` with your HuggingFace API token.

Usage

1. Start the application

```
streamlit run app.py
```

2. Open the browser and the required field in the interface
3. Press Submit

Example

1. Input a natural language query:

```
"Find all movies that have won more than 3 awards."
```

2. The system will generate the corresponding MongoDB query using Huggingface LLM and Langchain:

```
[
  { "$match": { "awards.wins": { "$gt": 3 } } },
  { "$project": { "title": 1, "_id": 0 } }
]
```

3. The generated query is executed on the MongoDB database, and the results are displayed in the pandas data frame.

Contributing

Contributions are welcome! Please open an issue or submit a pull request for any improvements or bug fixes.

Acknowledgements

- Huggingface Transformers <https://github.com/huggingface/transformers>
- Langchain <https://github.com/hwchase17/langchain>
- FAISS <https://github.com/facebookresearch/faiss>