

Hadith Search Chatbot – Report

M.Junaid 109 (4B)

Table of Contents

1. Introduction	1
2. Libraries Used	1
3. Functions and Workflow.....	1
4. Code Execution and Output	4
5. Screenshot of Output.....	4

1. Introduction

This Python-based Hadith Search Chatbot enables users to ask questions in natural language and receive relevant Hadith responses using semantic search. It uses data from the LK Hadith Corpus and processes the data using natural language processing and deep learning techniques to provide accurate search results.

2. Libraries Used

- pandas: For reading and manipulating Hadith data in CSV format.
- re: For text cleaning and regular expression-based preprocessing.
- sentence-transformers: For generating semantic embeddings of Hadith text.
- faiss: For indexing and performing fast similarity search on the embeddings.
- glob: For recursively loading multiple CSV files from subdirectories.

3. Functions and Workflow

- Data Loading:
 - All CSV files from the LK-Hadith-Corpus directory are loaded using glob.

```
# import glob
path = 'LK-Hadith-Corpus'
files = sorted(glob.glob(path + '/*/*/*.csv', recursive=True))

print(len(files))
```

335

- Relevant columns like English Hadith text and metadata are extracted.

```
all_hadith = []
for file in files:
    df = pd.read_csv(file, names=colnames, skiprows=1)
    # print(df.columns)
    # break
    if 'English_Hadith' in df.columns:
        df['Clean_Hadith'] = df['English_Hadith'].astype(str).apply(clean_text)
        all_hadith.extend(df[['Chapter_Number', 'Chapter_English', 'Section_Number',
                             'Section_English', 'Hadith_number', 'English_Hadith', 'Clean_Hadith', 'English_Grade']].values.tolist())
```

- Text Cleaning:

- Punctuation and extra whitespace are removed using a regex-based clean_text() function.

```
import re

def clean_text(text):
    text = text.lower()
    text = re.sub(r'^a-zA-Z0-9\s', '', text)
    text = re.sub(r'\s+', ' ', text)
    return text
```

- Embedding:

- The SentenceTransformer model 'all-MiniLM-L6-v2' is used to convert cleaned Hadith texts into dense semantic vectors.

```
model = SentenceTransformer('sentence-transformers/all-MiniLM-L6-v2')
```

✓ 4.9s

```
embeddings = model.encode(hadith_df['Clean_Hadith'].values)
```

- These vectors are saved to disk using NumPy for future use.

```
import numpy as np
embeddings = np.array(embeddings)
```

✓ 0.0s

```
np.save('hadith_embeddings.npy', embeddings)
```

- FAISS Indexing:

- A FAISS index is built using L2 (Euclidean) distance to enable fast similarity search on Hadith embeddings.

```
import faiss

dimensions = embeddings.shape[1]
faiss_index = faiss.IndexFlatL2(dimensions) # L2 Distance (Euclidean Distance)
```

✓ 0.8s

```
faiss_index.add(embeddings)
```

✓ 0.0s

Generate

+ Code

+ Markdown

```
faiss.write_index(faiss_index, 'faiss_index.faiss')
```

✓ 0.3s

- Querying Hadith:

- The user input is converted to an embedding and compared with the indexed Hadith embeddings.
- The top 5 closest Hadiths are retrieved and displayed based on similarity.

```
def get_similar_hadith(query, model, faiss_index, count=5):
    query_embedding = model.encode([query])
    distance, indices = faiss_index.search(query_embedding, count)

    results = []
    for i in range(count):
        hadith_text = hadith_df['English_Hadith'].iloc[indices[0][i]]
        score = distance[0][i]
        results.append({
            'Hadith Number': i + 1,
            'Distance': score,
            'Hadith': hadith_text
        })

    return results
```

4. Code Execution and Output

- The chatbot runs in the terminal and waits for the user to enter a question.
- It returns five Hadiths that are most relevant to the user's query, along with similarity scores.
- The chatbot keeps running until the user types 'exit'.

5. Screenshot of Output

```
hadith_Chatbot()
✓ 13.4s Python
```

Welcome to the Hadith Chatbot!
Query: how many prayers

Hadith 1
Distance: 0.6540
It was narrated that Ibn 'Abbas said: "Your Prophet (ﷺ) was enjoined to do fifty prayers but he returned to your Lord 1

Hadith 2
Distance: 0.7181
Abu Huraira reported Allah's Messenger (ﷺ) as saying: Prayer said in a congregation is equivalent to twenty-five (praye

Hadith 3
Distance: 0.7262
It was narrated from Abu Hurairah that the Messenger of Allah (ﷺ) said: "Praying in congregation is twenty-five portior

Hadith 4
Distance: 0.7308
It was narrated from Abu Sa'eed Al-Khudri that: The Prophet (ﷺ) said: "If one of you does not know whether he prayed th

Hadith 5
Distance: 0.7490
Narrated `Abdullah bin `Umar: Allah's Messenger (ﷺ) said, "The prayer in congregation is twenty seven times superior to