Project 1

Israel-Hamas War Question Answering System

Objective

The goal of this project is to develop a basic Question Answering (QA) system that can answer questions about the Israel-Hamas war using a dataset of news articles. The system should filter relevant articles, clean the data, and provide answers to specific questions about the conflict.

Dataset

The dataset is a JSON file with 37,000 news articles from various sources, covering a range of topics, including the Israel-Hamas war. The articles span from October 2023 to March 2024.

Approach

The solution can be divided into the following steps:

1. **Loading and Exploring Data**: Read the JSON file and load the data into a pandas Data Frame.
2. **Filtering Relevant Articles**: Filter articles that specifically mention "Israel" and "Hamas" in their content.
3. **Data Cleaning**: Clean the text data by removing unnecessary spaces, punctuation, and special characters.
4. **Setting Up a QA Model**: Utilize a pre-trained BERT model from Hugging Face's transformers library for the QA task.
5. **Implementing the QA System**: Create functions to process the input question and return relevant answers from the context provided by the filtered articles.

Two main functions are defined:

- `answer_question:` Uses the QA pipeline to get the top 3 answers for a given question and context.
- `ask_question:` Aggregates the cleaned article bodies into a single context and passes the question to `answer_question`.

Some Sample Output

1:-

```python
question = "what happened at the Shifa Hospital?"
answers = ask_question(question, israel_hamas_articles)
for i, answer in enumerate(answers, 1):
    print(f"Q: {question}\nA{i}: {answer}")
```

[5]

```
/usr/local/lib/python3.10/dist-packages/transformers/pipelines/question_answering.py:326: User
  warnings.warn("topk parameter is deprecated, use top_k instead", UserWarning)
Q: What happened at the Al-Shifa Hospital?
A1: bombed
Q: What happened at the Al-Shifa Hospital?
A2: Hospitals are bombed
Q: What happened at the Al-Shifa Hospital?
A3: Hospitals are bombed
```

⊗ 0 ⚠ 2   📡 0   ✓ AWS: AWS Builder ID   ▷ Amazon Q   🌐 Connected to Discord

2:-

```python
def ask_question(question, articles_df):
    context = ' '.join(articles_df['cleanedArticleBody'].tolist())
    answers = answer_question(question, context)
    return answers

question = "Describe the events that took place at the Shifa Hospital during the Israel-Hamas conflict"
answers = ask_question(question, israel_hamas_articles)
for i, answer in enumerate(answers, 1):
    print(f"Q: {question}\nA{i}: {answer}")
```

[6]

```
Q: Can you describe the events that took place at the Al-Shifa Hospital during the Israel-Hamas conflict?
A1: thousands dead and more than 1 million displaced in Gaza
Q: Can you describe the events that took place at the Al-Shifa Hospital during the Israel-Hamas conflict?
A2: They have stopped the supply of food water and fuel to Gaza Hospitals are bombed
Q: Can you describe the events that took place at the Al-Shifa Hospital during the Israel-Hamas conflict?
A3: Hospitals are bombed
```