

Problem Description

Attached with document is Json file containing 37000 news articles from different sources reporting about the Israel Hamas war. It contains articles between October 2023 to March 2024. Not all articles talk about the Israel-Hamas war, the Json file also contain some spurious articles talking about different topics. Also, since the data has been scrapped from the web, the articles may contain noise such as punctuation, special characters, and irrelevant symbols.

Q1. Create a rudimentary Question Answering system that can answer questions related to the Israel Hamas war.

1. You can use any open-source LLM (LLAMA, T5, BERT), library (langchain (<https://www.langchain.com/>), LLamaIndex (<https://www.llamaindex.ai/>)) or algorithm (BM25, Maximal Marginal Relevance) that you want.
2. Also, you can augment the existing articles by extracting additional information from Wikipedia or DBpedia

You will be judged based on the relevancy and readability of the answers generated.

Example question can be “**What happened at the Al-Shifa Hospital**”?

Q2. Create a timeline summarization plot or list, that shows the major events of the Israel-Hamas war that had occurred during the time-period mentioned in the Json file.

A timeline summarization basically details the major events that had taken place during a particular time-period. Please refer to this medium article for further information (<https://towardsdatascience.com/natural-language-processing-event-extraction-f20d634661d3>)

An example of a time timeline summarization list is as follows (excerpt from Wikipedia):

7 October – 27 October 2023: Initial attacks

28 October – 23 November 2023: Invasion of the Gaza Strip

24 November 2023 – 11 January 2024: First ceasefire

12 January 2024 – 6 May 2024: Yemen airstrikes

7 May 2024 – present: Rafah offensive

Attempt anyone one of the two questions that you like. Fork this repository and create a branch. Once you have completed your code push the branch into the repo for me to evaluate.

You will need to submit the following items:

1. **Ipython notebook or python script**
2. **requirement.txt** containing the python packages needed to run the code
3. And a document explaining your thinking process and code.
4. Also, a **Contact Me** file containing your contact information which we will use to contact you.