

### **Understanding the Problem Statement:**

We are tasked with creating a question-answering system based on a dataset of news articles, focusing on topics related to Israel, Hamas, Gaza, Palestine, war, and conflict. We have to create a Question-answer system which gives answers to the particular users query.

### **Analysing the Dataset:**

We have a JSON file containing news articles. Each article has fields like 'articleBody', 'dateModified', 'scrapedDate', 'source', and 'title'. We'll need to preprocess this data to make it suitable for our question-answering system.

### **Preprocessing the Data:**

We then preprocessed the text by converting it to lowercase, removing unnecessary characters, and tokenising it. This will make it easier to work with the text data later on.

### **Filtering Relevant Articles:**

We then filtered out articles that were relevant to our topics of interest. This can be done by checking if certain keywords (e.g., 'Israel', 'Hamas') are present in the article text.

### **Implementing BM25 Retrieval:**

After carefully analysing the data and after preprocessing I have decided to BM25 retrieval algorithm. It is used for search queries and BM25 takes into account both term frequency (TF) and document length normalization to determine the relevance of a document to a given query.

So I initialized BM25 using tokenized articles.

### **Implementing T5 Question-Answering Model:**

Then I used T5 model fine-tuned for question-answering tasks to generate answers based on the user query and the retrieved relevant articles.

### **Putting It All Together:**

At last, I created a loop where the user can input questions, and the system will generate answers based on the input. The loop will continue until the user decides to exit the system.