After reviewing the description and analyzing the dataset, I decided to use BM25 for building a Question-answer system. BM25 is designed for search queries, which makes it suitable for this task. Additionally, the dataset is in JSON format and consists of news articles, so I imported specific libraries to facilitate the process. These included **json** for handling JSON files, **re** for regular expressions, **tqdm** for progress bars**, rank_bm25** for BM25 search, and **transformers** for utilizing the T5 model.

Following the library imports, I loaded the dataset and preprocessed it using a clean article function. This function was responsible for removing special characters, punctuation, and converting text to lowercase. It also filtered out articles containing specific keywords such as "Israel" or "Hamas".

After data preprocessing and filtering, the articles were tokenized to prepare them for BM25 search. I initialized BM25 using the tokenized articles. I then defined functions for both questions and answers.

To retrieve the most relevant articles based on a given question, I created a function called **retrieve_articles**. To generate an answer to a given question, I used the **answer_question** function. Furthermore, I developed a function specifically for questions related to Israel and Hamas, named **answer_israel_hamas_question**.

To enable user input for asking questions, the script enters a loop where it continuously prompts the user to input a question. Upon receiving a question, it provides an answer and then prompts for another question. To exit this loop, users can simply type "exit," which will break the loop.