IR Project Report

1) **First** imported the necessary libraries, and read the vaswani dataframe

2) **Preprocessing:** I performed preprocessing on the column named "Text". I cleaned it, applied stemming, and removed stop words. Then, I placed the processed text into a new column in the data called "processed_text".

3) **Indexing:** I generated the inverted index for the terms I have in the data, and for each term, maintained a list of document IDs where the term appears along with the frequency of occurrence.

4) **Query Processing:** I performed preprocessing steps on the query and created two functions from scratch: the first one to Identify relevant documents by leveraging the inverted index. And the second one to retrieve documents that contain all the terms from the query. After that, I ranked the documents based on TF-IDF.

5) **Query Expansion:** Used the RM3 to expand the query , after that used ELMO.

6) **User Interface:** I created a search engine using Flask, where you can input any query, and it will return the relevant documents.

7) **Evaluation:** I conducted an evaluation of the search engine.