Shreayan Chaudhary
March 7, 2023
Homework #2b Information Retrieval 601.666

PART C:

Extension 10:

The first extension I have implemented is extension number 10, where I can accept queries directly from keyboard, and retrieve the top 20 most relevant documents. The user has the option to quit or continue with another query, and then they can choose to enter queries into the categories of author, title, keyword, and abstract, and they can continue this until they choose to stop. Once the user provides all their queries, they will be converted into the document class with doc_id, author, title, abstract and keywords and the relevance ranking will be calculated using these queries. Given below is a sample prompt for the input from user:

```
(base) PS C:\NonOSFiles\BlueJayCodes\IRWA\hw2> python .\hw2.py
Enter the query you would like to look up: tell me about the weather
forecast
Would you like to enter more queries? Y/N: y
Enter the query you would like to look up: california wildfires
Would you like to enter more queries? Y/N: y
Enter the query you would like to look up: champions league winner
team soccer
Would you like to enter more queries? Y/N: n
```

Since it's an interactive extension, I manually added the output on the terminal to file input_from_user_relevance_ranking.tsv.

Extension 2:

The next extension I have implemented is extension number 2. I created a dataset of size 100 documents called news.raw for different news and articles, and 10 queries in news_query.raw regarding information about news. I have also listed all the relevant documents in news_query.rels. I have also provided relevant judgment for these queries. Please refer to file news_top_20_docs.tsv and new_top_10_docs.tsv to view the relevance judgment for the queries and the news data that I have created.