**Name: Fabliha Faiza Maliha**

**ID : 500910780**

**Assignment 2**

There are four files in this program: **search.py, invert.py, eval.py, test.py**. For implementations, I used Python as a programming language. The invert.py processes the given cacm.all data based on given stopwords and creates the dictionary and posting list. Then I implemented the test function which contains the *test()* function as a driver function. The *test()* function first processes the query.txt and qrels.txt files based on the user inputs. Then it calls the *search\_query()* function and search in the saved dictionary to check if it has appeared in any documents. The *search\_query() function returns top-30 relevent documents ID and* printing the *document ID* with *title, authors, and score.*

For this assignment, I am calculating the score of all the relevant documents and printing the top 30 relevant documents. I have used *cosine similarity* technique to calculate the scores and I have also predefined the value as 30 in the **test.py** file. For the weighting scheme for document and queries, I have used the following variation of tf-idf weighting scheme respectively:

**TF** : 1 + , **IDF** : 1+

**weight\_of \_doc** : tf \* idf

**weight\_of\_query:** tf \* idf

After calculating the MAP and R-P values for all the queries, I get the approximate value :

***average mAP : 0.145***

***average R-Precision : 0.352***

which means that my program found 35% of the total relevant documents within the number of results equal to the number of relevant documents.

**Screenshots :**





