CPS842 Fall2021
Assignment 2 Report

Name: Tusaif Azmat Student#500660278

Instructions:

There are three files to achieve the tasks for the assignment 2.

1- Invert.py file
2- Search.py file (user interface also)
3- Eval.py file
4- Search_ui.py (user interface for user search query)

1. Invert.py file

This program file **invert** constructs the inverted index from CACM file. This program creates the postings for each term in the list with document ID, and document position in the actual document. It makes the more efficient retrieval process.

2. Search.py file

This program implements a vector space model for information retrieval system. It uses TF-IDF weighting schemes to normalize document length and term frequency. The corpus used is a collection of papers and their abstracts (cacm.all file). It auto-calculates the evaluation metrics.
You could execute this program for user interface as well. In this program user could interactively enter a query, and the program will return all the relevant results. For each result, the ranking order (e.g. 1, 2, 3), the document title and the author names are displayed.

3. Eval.py file

This program evaluates the performance of the IR system. The program **eval** takes input to this program as two files, query.text and qrels.text from CACM. The program go through all the queries in query.text, for each query, get all the relevant results from the retrieval system (by running **search**), compare the results with the actual user judgment from qrels.text, and then calculate the mean average precision (MAP) and R-Precision values. The final output will be the average MAP and R-Precision values over all queries.

4. Search_ui.py

By running this program users could interactively enter a query, and the program will return all the relevant results. For each result, the ranking order (e.g. 1, 2, 3), the document title and the author names should be displayed. This program basically uses the output from **search** program.

Input Files to the information retrieval system.

cacm.all	text of documents
stopwords.txt	stop words that can be ignored by the program during the inverted index creation
posting.txt	inverted index file.
query.text	Original text of the query
qrels.text	relation giving qid did 0 0 to indicate document did is relevant to query qid

Program execution instructions.

1. Execute the 'invert.py' that will create the inverted index file. It will ask user to provide input towards using or to ignore stop words and use the stemmer.

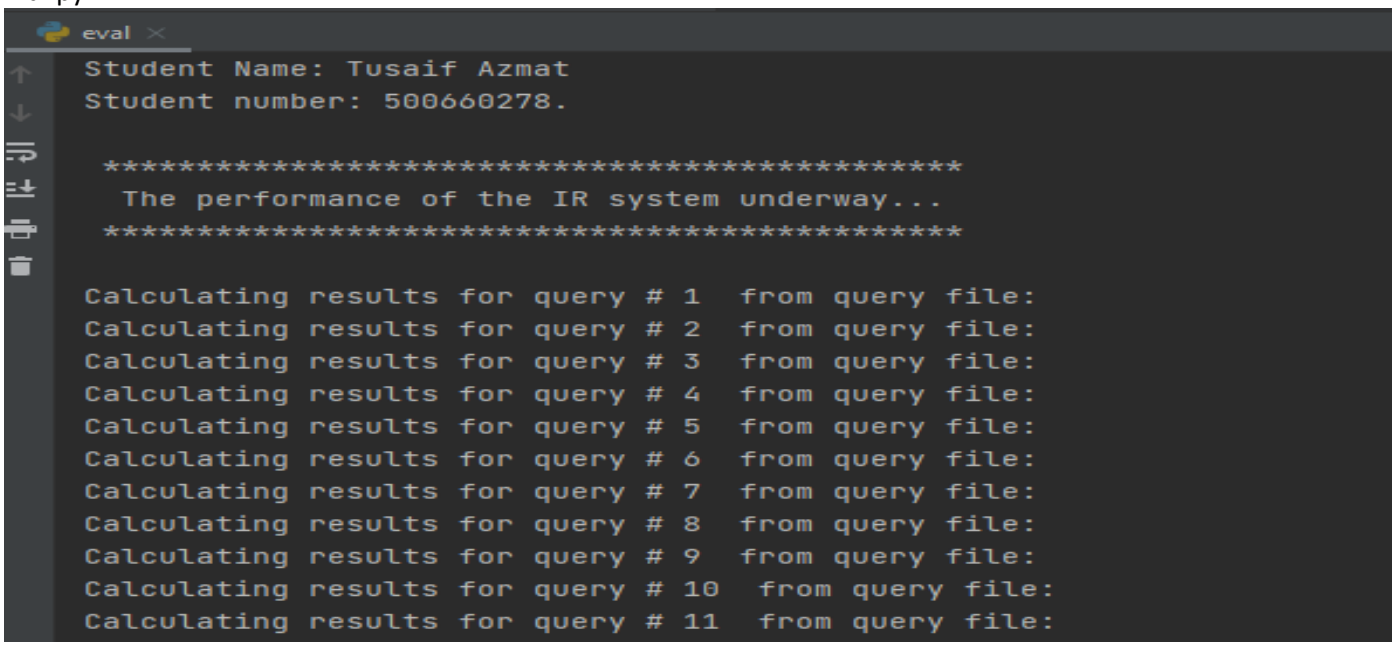
2. Execute the 'search.py' that will provided user with query search option. This program provides with a query of user's choice on the document collection and return the relevant results ordered by rank, document id and document position.

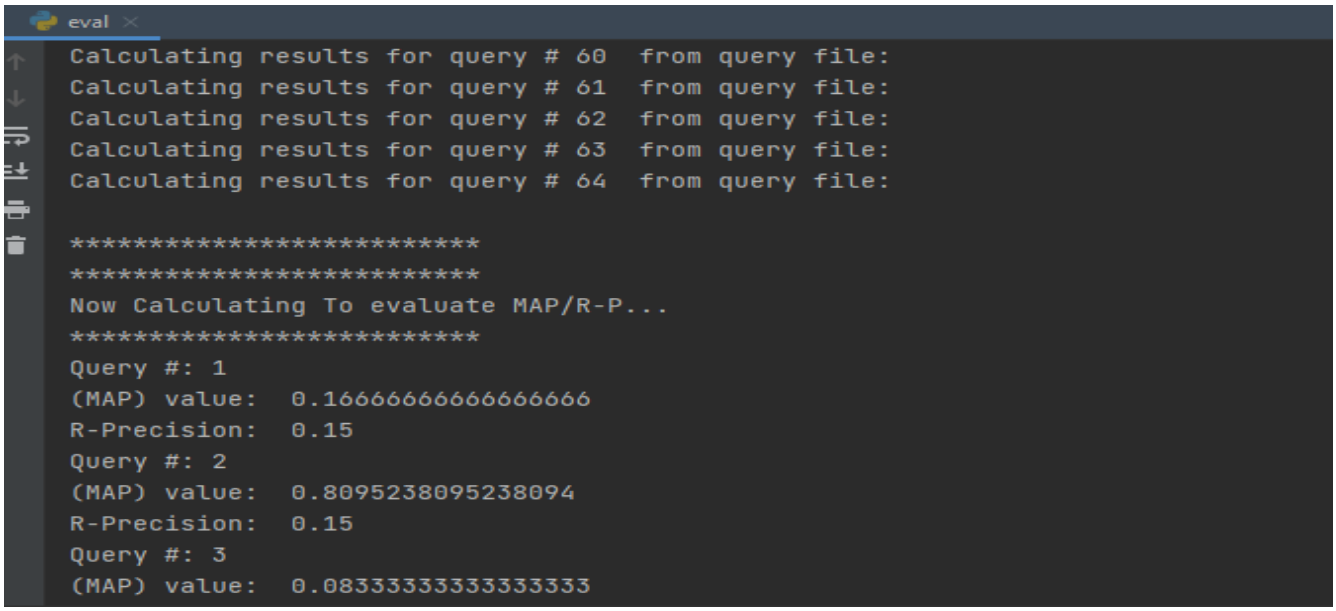
3. Execute the 'eval.py' to run the automatic evaluation program. This program uses the queries in query.text, and evaluates the search program against the lists of relevant docs given in qrels.text to give Mean Average Precision (MAP) and R-Precision values.

4. Execute the 'search_ui.py' and users could interactively enter a query, and the program will return all the relevant results. For each result, the ranking order (e.g. 1, 2, 3), the document title and the author names should be displayed. This program basically uses the output from **search** program.

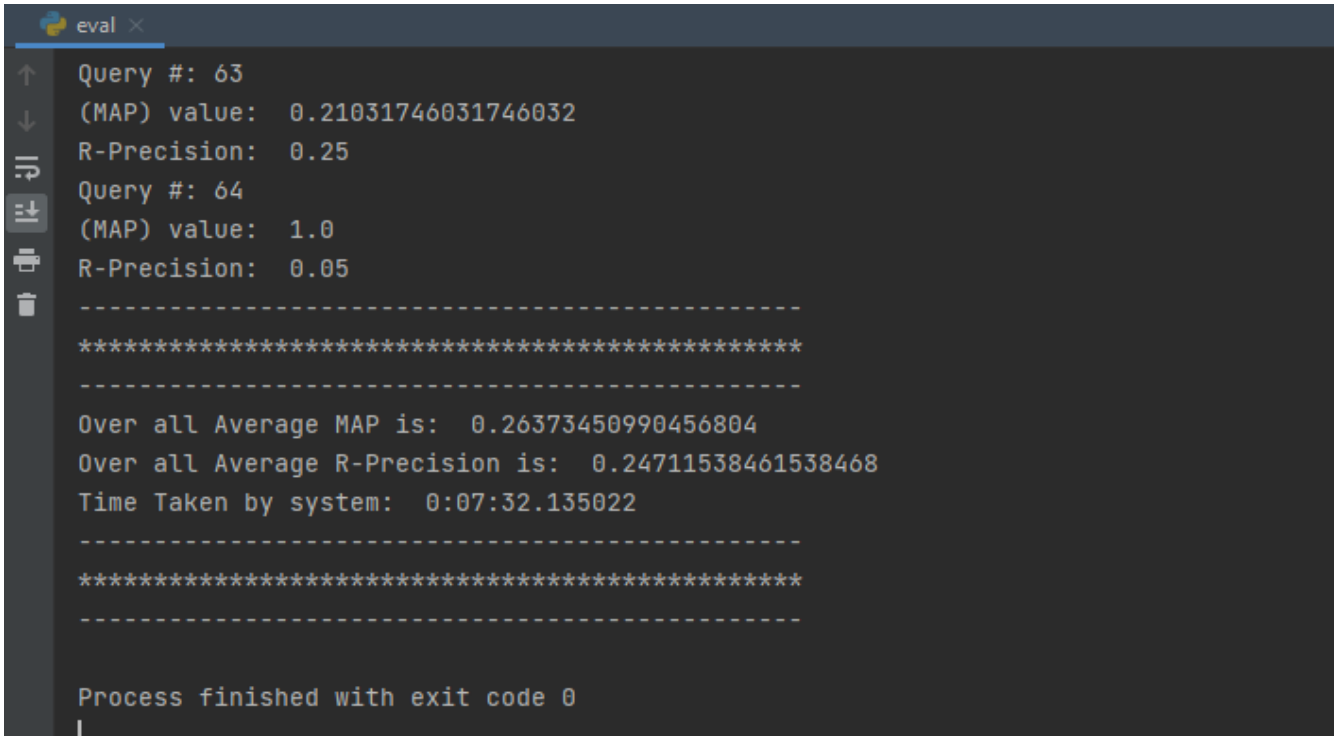
Screen shots:

Eval.py




eval
Calculating results for query # 60 from query file:
Calculating results for query # 61 from query file:
Calculating results for query # 62 from query file:
Calculating results for query # 63 from query file:
Calculating results for query # 64 from query file:
****
****
Now Calculating To evaluate MAP/R-P...
****
Query #: 1
(MAP) value: 0.16666666666666666
R-Precision: 0.15
Query #: 2
(MAP) value: 0.8095238095238094
R-Precision: 0.15
Query #: 3
(MAP) value: 0.08333333333333333

...


eval
Query #: 63
(MAP) value: 0.21031746031746032
R-Precision: 0.25
Query #: 64
(MAP) value: 1.0
R-Precision: 0.05

****

Over all Average MAP is: 0.26373450990456804
Over all Average R-Precision is: 0.24711538461538468
Time Taken by system: 0:07:32.135022

****

Process finished with exit code 0

Search.py / User Interface: