

Homework 3- Indexing and Retrieval

Language: Python

Version: 3.5

Tool: PyCharm

Libraries:

1. Math
2. json

Source files:

1. indexer.py
2. bm25.py

Steps to run the code:

1. Execute the indexer file directly without parameter on command line and follow the prompts
 - The execution will prompt for corpus. Provide the corpus file name as tccorpus.txt
 - If index.out already present, the program will delete the file and print message asking to rerun the code
 - If index.out is not present, the program will generate the index.out file and print a message for successful file creation

2. Execute the file bm25.py directly without parameter on command line

- The program will prompt for index file. Give input as index.out
- The program will then prompt for queries file. Give input as queries.txt
- The program will then prompt for maximum number of documents to be evaluated.
- Once the program finishes it will display the top 100 documents for 7 queries on standard output.

Note: The input files to the program should be in same directory as the program is.