CPS842 Fall2021
Assignment 1 Report

Name: Tusaif Azmat Student#500660278

INSTRUCTIONS ON RUNNING THE PROGRAMS

Step 1: User must run **Invert.py** first to create the **dictionary.txt and postings.txt** file and document_word.txt (this is the help txt file).

Step 2: The program takes the input of the cacm collection and then begins to read through the file line by line. For each line read, it is determined whether the information is required to be stored by the program or not. If it is required, it then begins to go through reach word and determine if the word should be stored in the dictionary and posting list.

Step 3: The program prompts user to enter if they want to use stop word option to include by turning it “on” or “off”. If stop words are being used certain words would be skipped.

Step 4: The program prompts user to enter if they want stemming option to include by turning it “on” or “off”.

Step 5: After the step 3 inputs, the program Invert.py is terminated, then the user needs to runs Test.py file.

Step 6: Test.py asks the user to enter the query work to search. After the query is entered in the prompt the program start to look from the three file created in the first three steps. If the user wants to terminate the search should type “ZZEND”.

Step 7: The program looks for the query word in the dictionary file and in for some reason it’s not found it will output word not found and asks user for input another word.

Step 8: If the query word is found in the dictionary. The frequency value that is associated with the query word will be printed. The program takes the query term to get the correct posting index by iterate through the value associated with the key that is the query term.

Step 9: The program use the posting list to print out the doc id, doc title, frequency of term in doc and also its position of appearance in doc. The program shows the terms’ first appearance in the document. The third file will provide the 10 words surrounding it for context. It also prints the query time and the average time in case of more than one query.

Step 10: Once the user types “ZZEND” the program terminates.

Screen Shots:

Invert.py

```
Run:  invert01
Do you want to use stemming? [Yes/No]yes
Would you like to remove stop words? (yes/no): yes

Process finished with exit code 0
```

Test.py

```
Run:  test01
Enter your term to search below or enter ZZEND to end the program.
Enter a Single Term to Search:stress
Document frequency : 10
Doc ID: 71 .
 Title: propos feasibl program system .
 Frequency in doc  1  in position(s) 67
program construct program system should stress discoveri sourc program statement mani
Doc ID: 1048 .
 Title: approxim solut axial symmetr problem .
 Frequency in doc  1  in position(s) 18
theori heat ideal fluid flow stress concentr theori reduc under assumpt
Doc ID: 1135 .
 Title: gener business-ori languag base decis express .
 Frequency in doc  1  in position(s) 99
convent mechan languag design philosophi stress structur analysi class process repres
Doc ID: 1392 .
 Title: experi formac algorithm design .
 Frequency in doc  1  in position(s) 51
formac implement present attempt made stress principl idea gener relev design
Doc ID: 1542 .
```

```
Run: test01
storag particular processor develop paper stress simultan oper within microinstruct adder
Doc ID: 2314 .
 Title: requir advanc program system list process .
 Frequency in doc  1  in position(s) 134
data form storag manag extens stress dualiti data retriev function evalu
Doc ID: 2765 .
 Title: analysi perform invert data base structur .
 Frequency in doc  1  in position(s) 16
system hierarch level level framework stress invert data base file organ
Doc ID: 2922 .
 Title: two-level control structur nondeterminist program .
 Frequency in doc  1  in position(s) 50
recogn these two level discuss stress structur manag choic level free
0.12499690055847168
Enter a Single Term to search:hello
hello is Not found in the Document
0.0
Enter a Single Term to search:go
Document frequency : 11
Doc ID: 321 .
 Title: algol 60 confidenti .
 Frequency in doc  1  in position(s) 65
other compil languag write assign go statement etc inde lot unnecessari
Doc ID: 1135 .
Problems  Run  Terminal  Python Packages  Python Console
```

```
Run: test01
 Frequency in doc  1  in position(s) 24
cycl node detach until new can enter acycl process import certain
Doc ID: 3172 .
 Title: algorithm plan collision-fre path among polyhedr obstacl .
 Frequency in doc  1  in position(s) 59
which indic each vertex transform obstacl which other vertic can reach safe
Doc ID: 3177 .
 Title: share secret .
 Frequency in doc  1  in position(s) 40
kei manag scheme cryptograph system can function secur reliabl even misfortun
Doc ID: 3188 .
 Title: semiot program languag .
 Frequency in doc  1  in position(s) 30
necessari cover all them on can howev project most aspect into
0.37497854232788086
Enter a Single Term to search:tusaif
tusaif is Not found in the Document
0.0
Enter a Single Term to search:ZZEND
End of the search.
Average query search time is:  0.10311770439147949

Process finished with exit code 0
```