Documentation for 1st part and 2nd Part (using grep and python)

-Hemanth Reddy Vennapusa

-15CS10051

First install perl by running the command -> `sudo apt install perl`

*****Preprocessing on documents -

```
perl -pi -e 's/\n/ /g' *
perl -pi -e 's/\r/ /g' *
```

the following two commands were used to replace all new lines and return carriage to space to help use grep

Document name format - 'GX[0-9]\{3\}\-[0-9]\{2\}\-[0-9]\{7,8\}'

if the query is 'W1 W2 W3 W4 W5' the following grep command can be issued to search in all documents and output the file names

```
grep W1 * -i | grep W2 -i | grep W3 -i | grep W4 -i | grep W5 -i | grep
'GX[0-9]\{3\}\-[0-9]\{2\}\-[0-9]\{7,8\}' -o
```

****The `output.txt,query.txt,alldocs(folder)`'s PATH is default been set to as in the programs folder

      but if its different we can change it in the global variables benchm , queries ,loc in starting of the source code .

      No of queries is also been defaulted to 82 in the global variable noquer.

The program first opens the query.txt file then for each query in the file it executes the grep command and appends the output to the
results.txt file and also counts the no of results for each query and total no of results for all queries.

It also calculates the time of execution for each query and stores it.

Then it opens the output.txt and for each query it maintains a set containing the relevant document ids

Then it opens the `results.txt` and compares the output and the results to calculate the precision and recall

The result is printed in the `performance.txt` file

To run the program make sure the directory of program is as follows

```
current dir /
```

```
-> searcher.cpp
-> makefile
-> alldocs/        (folder)
-> output.txt
-> query.txt
-> presenter.cpp  (this one just formats the output of grep to results.txt)
```

if not change the path variables as stated above

After above step open a terminal and type -> `make clean`
                                                               -> `make run`