Trisha Espejo

CSCI335

Project 1 Phase 3

Due: 5/11/21

Installation :

In this project there are 6 java files

file.java – will be use to storing all the filename and content use for inverted index

HtmlProperties.java - use to save all html file name and raw content

Project2.java - main program where main method is located

SearchEngine.java - where all the work is performed

createStem.java - change the word to its root word

stemmer.java - performs the stemming

resultGUI.java // class that creates gui

Also stemmer.java is not my code it came from the website: <https://tartarus.org/martin/PorterStemmer/java.txt>

there are 4 txt needed to run the code

invertIndex.txt - outputs the inverted index

query.txt - query you want to look up

result.txt - result of query result

stopwords.txt - list of all stopwords

To generate html file

copy source code of webpage and put it in folder title it html

So in running the file you must have project folder

src code is inside the project folder with all the java code in source folder

you also need to have html folder inside the project folder with all the html webpage

you also need to have invertindex.txt, query.txt, resut.txt and stopword.txt inside the project folder

There would be a total 7 txt file in the project submission

invertIndex.txt - outputs the inverted index of without stemming

query.txt - query you want to look up

result.txt - result of query result of without stemming

stopwords.txt - list of all stopwords

invertIndexStem.txt - outputs the inverted index of with stemming

resultStem.txt - result of query result of with stemming

so you need to add all the text file inside the project folder.

The inverted index.txt, result.txt can be empty

But the query.txt and the stopwords.txt need not to be empty

You can enter any word or sentences in each line of the query

The stopwords.txt is provided

You then need to insert html folder that is provided inside the project folder

It has list of all html

You can then run your program without trouble