Read Me File

The dataset I have used for this application is the time magazine articles with a total of 423 documents.

I built my application using java sockets for the client side and server-side features. After running the main method inside the main class go to a browser (**preferably Mozilla Firefox,** didn't work in chrome) and use the url:

"http://localhost:3000/".

There are 4 method calls in the main method out of which 3 are commented:

Utilities.createDocs: This method creates a separate file for each document and stores it in a directory. I ran this method only once for pre-processing purposes. This method is commented out

Utilities.invertedIndex: This method builds the inverted index and stores it in a file invertedIndex.txt in the project root folder. This method is commented out after running it once for the first time. It creates the index using tf x idf values for term weights. The structure of classes and algorithm used were taken from the implementation notes taught in the class. This method sends all the documents to the Stemmer class for stemming which uses porter's algorithm

Utilities.getInvertedIndex: This method retrieves the index from the file which was created by using the above method once. This is the only method that gets executed after running the program

Utilities.relevanceAssesment: This method is also commented out but if you uncomment this method and run the application first it retrieves the index and then runs all queries from the TIME.QUE file and compares the results with TIME.REL file and evaluates the system using precision and recall, it will create two files in the project root folder. After assessing it will run the server

- relevanceAssesment.txt: this file has the query numbers and then precision and recall in the following pattern: 1 0.75 0.8571428571428571
- myRetrievedResults.txt: this file has the query number and the numbers of documents that my system has retrieved for that query and its format is:
  1 308 257 370 323 326 304 268 334

If we run the application with uncommenting only Utilities.getInvertedIndex method, then it retrieves the already built index and starts the server and in the console, it displays the message: **Retrieval System is starting up, listening at port 3000.**

Once we see the above message we can go to Mozilla firefox and type in **"http://localhost:3000/".** In the address bar and we will be directed to a page that has a search bar and search button. Type in the query in the bar and hit search and you will be redirected to the results page where it shows the hyperlinks of the retrieved documents and clicking on them will take you to the document itself.

As soon as the query comes in from the browser to server, it is passed to the processQuery method in the Utilities class. This method stems the query and finds the tokens and then calculates the cosine similarity between the query and documents and then ranks the accordingly.

Outside resources: porter's algorithm for stemming:
https://tartarus.org/martin/PorterStemmer/java.txt

P. S: I ran the assessment method and the relevanceassesment.txt and myRetrievedresults.txt files have been created and I have commented the Utilities.relevanceAssessment method. If you want to recheck it hen delete those files and uncomment the Utilities.relevanceAssesement method run the application again.