CSC849 HW1
Chohee Kim - 915409942

Classes

- ● Value
- - Customized class that contains document frequency and postings list
  - ○ Data fields :
    - ■ dFreq:int - Document Frequency
    - ■ postingList:List<Integer> - Postings List
  - ○ Methods :
    - ■ Gettings and settings
- ● Parser
- - Parse the text file and has method that returns myDocumentTreeMap which maps from docId to a text that belongs to the unique docId.
  - ○ Data fields :
    - ■ textReader:Scanner - Scanner that reads given file.
    - ■ myDocumentTreeMap:TreeMap<Integer, String>
  - ○ Methods :
    - ■ getMyDocumentTreeSet()
    - ■ etc..
- ● Indexer
- - Tokenize the text for each unique documents and has a method that returns invertedIndex, which maps from a term to Value.
  - ○ Data fields :
    - ■ myDocumentTreeMap:TreeMap<Integer, String>
    - ■ invertedIndexList:TreeMap<String, Value>
    - ■ krovetzStemmer:KrovetzStemmer
  - ○ Methods :
    - ■ getInvertedIndexList()
    - ■ etx..

- ● Query
- - Query a given term and has a method that returns result of the search.
  - ○ Data fields:
    - ■ invertedIndexList:TreeMap<String, Value>
    - ■ searchTerms:String[]
  - ○ Methods:
    - ■ runQuery()
- ● SimpleSearchEngine
- - A class that contains main method.
  - ○ writeInvertedIndexListToFile(TreeMap<String, Value> invertedIndexList)
  - ○ runSearch(TreeMap<String, Value> invertedIndexList, TreeMap<Integer, String> myDocuments)

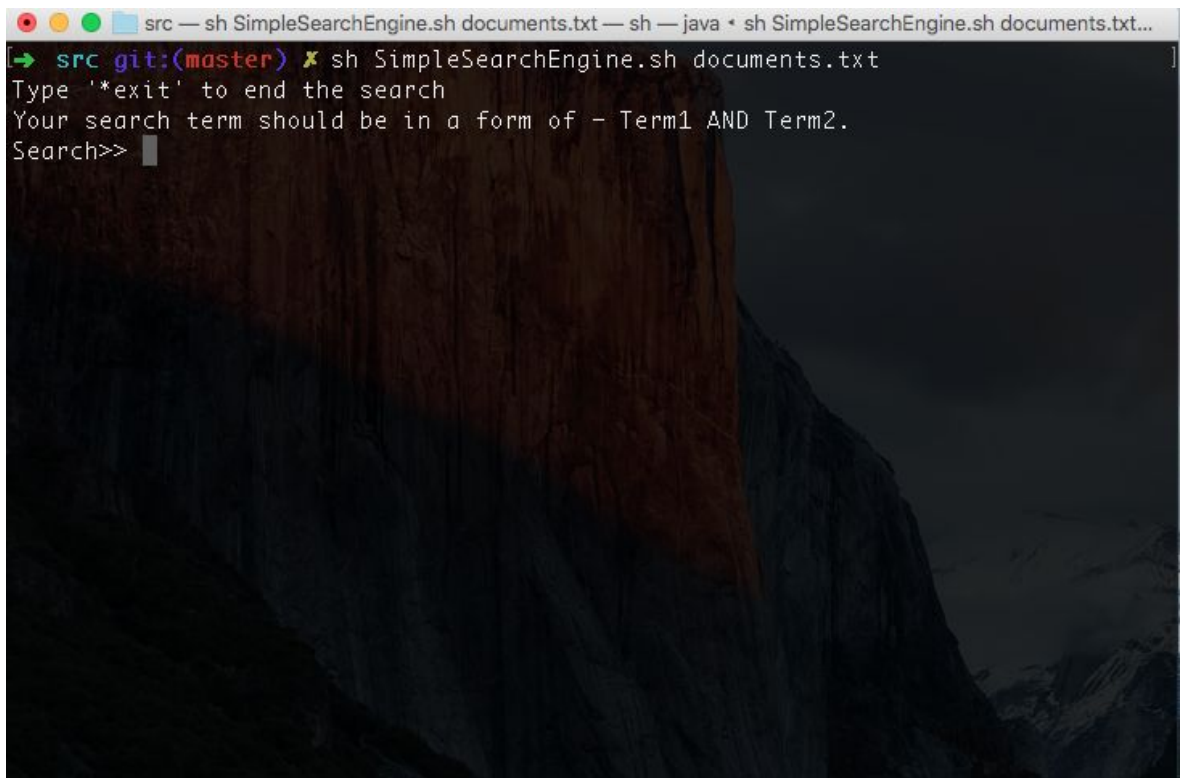## How to Run

1. Unzip the fold.
   a. The fold contains :
      i. Five .java files - Value, Parser, Query, Indexer, and SimpleSearchEngine
      ii. Kstem-3.4.jar
      iii. SimepleSearchEngine.sh
   b. In console, type to run the script
      sh SimpleSearchEngine.sh [textfile.txt]
      **The textfile must be in the same directory where all files are in

**e.g sh SimpleSearchEngine.sh documents.txt

c. The program runs, and type search term. To terminate the program, type *exit

   **Search term should be in a form of Term1 ANd Term2

        E.g. Google AND Nexus

          Google AND Asus

d. In the same directory, there will be two text files created :

   **You will also be able to see result on console, but text for each document

    i.   Inverted_index_file.txt - has inverted index list

    ii.   Result_file.txt - has result of the search