CPS842 Assignment 1 : Report

Algorithms and Data Structure for invert.java

The major algorithm used was Porter Stemming Algorithm to create the stemming option that the program can use when creating a stemmed dictionary and/or posting list.

The data structure primarily used was Arrays and Hash Maps, these specific data structures were used to tackle the issue that arises when creating an object, in this case term, that would hold its own document frequency, term frequency, document identification, and position in said document. The use of hash maps linked the word, as a key, and the term, as an object. Since hash maps does not have its own way of retaining all of its key in an easy to store and retrieving method, the incorporation of Array, specifically arraylists, to store the key words to retrieve the objects values after.

How to run : invert.java

To run the program as indeed there are 4 files needed for the first part, invert.java, cacm.all, stem.java, and term.java. To run the first part all 4 files need to be in the same file project space.

Executing first program: invert.java

When executing the first program it asks for user inputs to turn off/on stop word and stemming when creating the dictionary. To turn on/off stop word and/or stemming, in respective order, user must input either true(for yes) or false(for no) and other input will cause an Input mismatch and will need the program to restart

Incorrect execution:

```
Stop word: On/Off, enter true for on or false for off - not true
Exception in thread "main" java.util.InputMismatchException
        at java.base/java.util.Scanner.throwFor(Scanner.java:939)
        at java.base/java.util.Scanner.next(Scanner.java:1594)
        at java.base/java.util.Scanner.nextBoolean(Scanner.java:1893)
        at invertedIndex.requestInput(invertedIndex.java:39)
        at invertedIndex.main(invertedIndex.java:26)
```

Correct execution:

```
Stop word: On/Off, enter true for on or false for off - true
Stop word set to : true
Stemming: On/Off, enture true for on or false for off -false
Stemming set to : false
```

Executing second program: test.java

For this program make sure that all files are in the same workspace, especially the dictionary and posting file created by the first program. When running input any valid query, for invalid queries the program will return a system response of no query found.
To end the program enter ZZEND

Note: the query is case sensitive and queries must only be in lower case.
Query Example: april

```
Enter a query :
april
Query : april
Document Frequency : 2

Document ID : 146
Title : The Use of Computers in Engineering Classroom Instruction
Term Frequency : 1
Positions : [11]
Summary : April 29-30, the Computer Committee of the College of Engineering,

Document ID : 2593

Title : A Back-end Computer for Data Base Management
Term Frequency : 1
Positions : [42]
Summary : April 1971 Report) from a host computer, accesses the data

Query : april executed in milliseconds : 0.128
To end enter : ZZEND
```