

Assignment #3 (7%)

Array/ArrayList, Sorting, Input/Output

Requirements:

- Read the data from docset2020.txt as the input.
- In your main method file, process the text content in a set of documents and perform following tasks:
 - Normalize the upper and lower cases
 - Remove stopwords (ref: stop words.txt)
 - (extra 1%) Perform stemming (ref: Stemmer.java)
 - Build a dictionary term file docProcess.csv.
This file should contain all UNIQUE index terms rank in alphabetic order, their collection frequency (i.e., how many time this term appear in the whole collection)

Example

```
0001 account 1
0002 activ 1
0003 ad 3
.....
```

- Write a method **sortByFreq** that order the output by collection frequency.
If two terms have the same frequency, output by the alphabetic order.
You can rank the data by any type of sort (e.g. bubble/search/sequential/merge sort).
- Output the sorted data in another file docProcessOrdered.csv.
- Generate complete Javadoc (e.g. import statement, class, method, comments, etc.) for the class and place them in a folder named docs within the project.

Submission: Your Java project is named `yourStudentID_HW3`. Put all files in a folder and compressed it. Submit your assignment on eCourse2 under HW3. No other submissions will be graded and points will be deducted for late submission.

Academic dishonesty: You may not do work for another student nor may any student copy or plagiarize someone else's work. Severe penalties will be imposed on all parties involved.

Deadline: Thursday, December 24, 2020. (end of the day)