108-1 Assignment #3 (12%)
Array/ArrayList, Sorting, Input/Output, Exception Handling

Requirements

- Read the data from **`docset1.txt`** as the input.
- In your main method file, process the text content in a set of documents and perform following tasks:

  [Task 1] 6%
  - Normalize the upper and lower cases
  - Remove stopwords (ref: stop_words.txt)
  - (extra 2%) Perform stemming (ref: Stemmer.java)
  - Build a dictionary term file **`docProcessOutput.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
    1 account 1
    2 activ 1
    3 ad 3
  - You can rank the data by any type of sort (e.g. bubble/search/sequential/merge sort).
  - Include error/exception handling in your code to dealing with program errors

  [Task 2] 6%
  - Build a posting file **`posting.csv`**. This file also include a repeated set of information which indicate the document id that the term is in, and the term frequency (i.e., how many time this term appears in this document) of each appearance of this term in this document.
    Example
    1 account 1 Document Freq:[0, 0, 1, 0, 0, 0, 0, 0, 0, 0]+DocNO:[APW20010124.1256.0442]
    2 activ 1 Document Freq:[0, 0, 0, 0, 0, 0, 0, 0, 1, 0] + DocNO:[CNN20001019.1400.0201]
    3 ad 3 Document Freq:[1, 1, 0, 0, 0, 0, 0, 1, 0, 0] + DocNO:[APW20010120.0310.0146, APW20010120.2106.0704, CNN20001019.1400.0012]
  - (extra 2%) Generate the term frequency for each term in the corresponding document and generate a term frequency file **`tf.csv`**.
    Example
    1 account 1 Term Freq:[0, 0, 0.016, 0, 0, 0, 0, 0, 0, 0]
    2 activ 1 Term Freq:[0, 0, 0, 0, 0, 0, 0, 0, 0.018, 0]
    3 ad 3 Document Freq:[0.002, 0.004, 0, 0, 0, 0, 0, 0.002, 0, 0]

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: Tuesday, January 7, 2020. (end of the day)