

## Assignment #3 (7%)

### Array/ArrayList, Sorting, Input/Output, Exception handling

#### Requirements:

- Read the data from docset110 HW3.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 case words to lower case
  - (extra 1%) Remove stopwords (ref: stop words.txt)
  - Build a dictionary term file dictionaryTerm.csv.  
This file should contain all UNIQUE index terms rank in **alphabetic order**, their collection frequency (i.e., how many times this term appear in the whole collection)  
**Example**  
*0001 account 1*  
*0002 activ 1*  
*0003 ad 3*  
*.....*
  - Write a method **sortByARI** that sort the output documents by their ARI scores in descending order. If two documents have the same score, output by the DocID order. You can rank the data by any type of sort (e.g. bubble/search/sequential/merge sort).  
**Example**  
*0001 20010123 12*  
*0002 20010120 7*  
*0003 20010121 3*  
*.....*
- Then Output the sorted data in another file docSortByARI.csv.
- Perform exception handling as necessary.

**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, December 28, 2020. (end of the day)