

Assignment #4:

Array, Input/Output, & Exception handling

Requirements:

In document processing, the minimum preprocessing steps that need to go through include normalize the upper and lower cases and remove stopwords. You are required to perform following tasks

1. Read the file `docset1.txt` as input
2. Normalize the upper and lower cases of words in the input file
3. Remove stopwords in the input file according to the predefined list in `stop_words.txt`
4. Export/write processed DOCNO and TEXT to another file named `yourStudentID_output.xml`
5. Include error/exception handling in your code to dealing with program errors

Grading policy: There is a minimum deduction of 10 pts for any program that violates the public interface. There is a 5 pt deduction for missing/incomplete javadocs, and a minimum deduction of 20 pts for an incorrectly submitted project.

Submission: Your Eclipse project is named `yourStudentID_4`. The project is submitted as `yourStudentID_4.zip`. Submit via eCourse. No other submissions will be graded.

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: June 1st