BLG 233E DATA STRUCTURES AND LABORATORY

EXPERIMENT 12 – STL

IMPORTANT REMINDERS

1. It is not allowed to use USB sticks during the lab sessions.
2. You should unplug your ethernet cables during the lab sessions.
3. Any reference book or help material (C++) is allowed.

In this experiment, you are required to analyze a text file using sets and maps in STL. Two input files will be given. First file, namely "stopwords.txt", contains a list of stop words. These words will be added to a set. Second file, namely "input.txt", will contain a text to be analyzed. This file should be read word by word and each world should be checked if it exists in the stop word set or not. If it exists in this set, this word should be added to a frequency map and the counter of this word in the map should be incremented by one. Otherwise, the program should move on with reading the next word from the input file.

At the end of the text processing, an output which shows the frequency of each stop word in the input file should be printed to the screen in decreasing order. This means that the stop word with the highest occurrence frequency should be written to the screen first and the stop word with lowest frequency should be written last.

An example output can be given as follows.

and 100 %10
……………..
………………
……………..

In this example, assume that the input file has 1000 stop words. In this case, output means that "and" is repeated 100 times in the input file and its frequency is %10. Frequency is calculated as follows: "and" stop word is counted 100 times and there are 1000 stop words in the input file. So it is %10. Note that this example is imaginary. Given input file may cause different results.