03-60-254: Data Structures and Algorithms
Lab Assignment 4: Word Count III Using Heap

1 Due date

In your lab section during the week November 12/14. To be submitted 15 minutes before the end of your lab section. Any submission within the last 15 minutes before the end of the lab session will not be accepted and a zero mark will be given. Late assignments will be given a zero mark. Submissions by email will not be accepted.

2 Regulations

This assignment must be done individually. Any similarity found between your code/answers and another student's, or a carbon-copy of an answer found in the web will be considered cheating. For the part where coding is required, you must implement your own program.

3 Objective

The aim of this assignment is to understand Heap and HeapSort.

4 The problem specification

The problem is the same as in A2, i.e., to count the frequency of words in a text file using the sorting method. This time, the sorting method is heapSort, and you will be required to write more code. More specifically, your tasks are:

1. (2 marks) Sort the words using HeapSort algorithm. You need to write the code for the heapSort. **Here** is the template for the code, including the MergeSort so that you can have a comparison. Note that this time the template can not run–you need to provide the missing code first for the *sink* procedure.

2. (3 marks) in A2 we can only print out the largest frequency. Heap is partially sorted, hence easier to get/remove the top values. Please construct a heap for frequency counts, and printout the top 20 largest counts when the data size is 1 million lines. Here is the link for the 1 million data dblp1m. More specifically, you need to:

   - construct a heap for frequency counts;
   - implement removeMax() method;
   - invoke removeMax() 20 times to obtain the top counts

You can also use other programming languages of your choice. For the frequency heap, a hint is that you can write a Heap class like below:

```
public class HeapInt {
        private int[] heap;
        int n=0;

        //Heap constructor
        public HeapInt(int[] a){
                n = a.length;
                for (int k = n/2; k >= 1; k--)
                        sink(a, k,n);
                heap=a;
        }

        public int  removeMax(){
                //your code goes here
        }
        ...
}
```