

## EEL 5934/4930 Advanced Systems Programming

### Assignment 1

**Due: Friday, January 22<sup>nd</sup> by midnight**

In this assignment you are going to implement three programs: Mapper, Reducer, and Combiner.

**Mapper program** gets the input from a file and for each word in the file generates key value pairs in the form (word, 1). You can assume that the words are sorted according to the first letter but words that start with the same letter may not be sorted. As an example, the input may look like

course cap class culture class cap course course cap culture concurrency course cap class door  
date date date door date door door error error eraser error eraser eraser eraser eraser field  
floor field field field field group group group group group group gap group group group gap gap  
gap hole hole hole hole hole heat heat heat house heat hierarchy inconsistency ice ice ice ice  
inconsistency

and the correct output (on the standard output) would be

(course,1) (cap,1) (class,1) (culture,1) (class,1) (cap,1) (course,1) (course,1) (cap,1) (culture,1)  
(concurrency,1) (course,1) (cap,1) (class,1) (door,1) (date,1) (date,1) (date,1) (door,1) (date,1)  
(door,1) (door,1) (error,1) (error,1) (eraser,1) (error,1) (eraser,1) (eraser,1) (eraser,1) (eraser,1)  
(field,1) (floor,1) (field,1) (field,1) (field,1) (field,1) (group,1) (group,1) (group,1) (group,1)  
(group,1) (group,1) (gap,1) (group,1) (group,1) (group,1) (gap,1) (gap,1) (gap,1) (hole,1) (hole,1)  
(hole,1) (hole,1) (hole,1) (heat,1) (heat,1) (heat,1) (house,1) (heat,1) (hierarchy,1)  
(inconsistency,1) (ice,1) (ice,1) (ice,1) (ice,1) (inconsistency,1)

Note that Mapper program should output (word,1) key-value pair as soon as it reads the word. Also, each key-value pair should be output on a separate line (although shown in a condensed form above.)

**Reducer program** gets the (word, 1) key-value pairs from the standard input as shown above and generates (word, total) key-value pairs on the standard output as follows:

(course,4)  
(cap,4)  
(class,3)  
(culture,2)  
(concurrency,1)  
(door,4)  
(date,4)  
(error,3)

(eraser,5)  
(field,5)  
(floor,1)  
(group,9)  
(gap,4)  
(hole,5)  
(heat,4)  
(house,1)  
(hierarchy,1)  
(inconsistency,2)  
(ice,4)

Note that Reducer displays the key-value pair as soon as it realizes that there won't be any more of that word, i.e., when it sees a word that starts with a different letter. You can assume that Reducer's input has a single key-value pair per line.

*Make sure that Reducer detects the end of file character while reading from the standard input.*

**Combiner program** gets its input from a file and for each word in the file generates the key-value pair (word, count). Each key-value pair is output on a separate line. You should use fork, exec, and pipe system calls in the Combiner and reuse the Mapper and Reducer programs.

**Important Note:** Implementations of Combiner programs that do not make effective use of these system calls will not get any credit.

**Submission:** You should submit your source code for the three programs along with a Readme and preferably a Makefile on CANVAS.