ADVANCED SYSTEMS PROGRAMMING
ASSIGNMENT - 1

DETAILS:

Name : Sai Vishnu Teja Vempali
UFID : 16141381
email : vishnu24@ufl.edu

INTRODUCTION :

The assignment consists of 3 C programs: Mapper, Reducer and Combiner

Mapper:

Mapper reads input from the file that was sent to it as an argument and generates key value pairs of the form **(word,1)** for each word in the input file.

Reducer:

Reducer reads the input from standard input(which is modified as the read end of the pipe). It reads the (word,1) pairs and generates key value pairs of the form **(word,total)** to standard output.

Combiner:

Combiner reads input from a file and generates key value pairs of the form **(word,count)** for each word in the file to the standard output.

IMPLEMENTATION:

Mapper:

Mapper reads input from the file that was sent to it as an argument from the combiner and generates key-value pairs of the form (word,1) and writes them to the write end of the pipe.

Reducer:

Reducer reads input from the read end of the pipe and it uses 2D array “args” to store the words and a 1D array “count ” to maintain the count of each word. Word stored at location pointed by args[i] has its corresponding word count value stored at location pointed by count[i].

When a word starting with a new letter is encountered, all the words in the args are printed in the (word,total_count) format and the args and count array memory id freed and fresh memory is created. Thus, Reducer outputs the key-value pair immediately when it sees a word that starts with a different letter.

To use memory efficiently, the reducer frees the memory allocated to a set of words starting with the same letter once they have been printed to the standard output.

Combiner:

It creates two child processes, one child process calls the mapper and the other child process calls reducer. This is achieved by using the execv() system call. The communication between the two child processes is achieved using pipe() and dup2() system calls.

The child2(calls reducer) process is made to wait until child1(calls mapper) is completed so that we do not encounter with error of reducer being executed before even the mapper is executed.

INPUT:

The input to the combiner program is given in the form of a text file(file.txt in this case).

OUTPUT:

key value pairs (word,count) is written to the standard output.

HOW TO RUN:

1. Download the zip file containing all the required files.

2. Extract the files.

3. Copy the Assignment1 directory to a convenient location.

4. Change directory to Assignment1. In the command prompt:

```
>  cd Assignment1

> make

> ./combiner file.txt
```

NOTE:

If any another file is to be given as a test case, in the command prompt, while running the executable please give as follows in the last step:

```
> ./combiner filename.txt
```