EEL 5733/4732 Advanced Systems Programming
Assignment 1
Due: Monday, January 22nd by midnight

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

Mapper program gets the input from a file that contains tuples in the form (userID, action, topic). userID is a 4-digit identification for the users of a social media site. Action can be one of the following letters: P for posting, L for liking, D for disliking, C for commenting, and S for sharing. Topic is a string of exactly 15 characters (shorter names will be padded with space). The mapper program processes these tuples to generate a weighted profile in the form (userID, topic, score) based on a set of rules, which are defined as P=50, L=20, D=-10, C=30, S=40. You can assume that the tuples are sorted according to userID field but tuples that belong to the same userID may not be sorted according to the topic. As an example, the input may look like

(1111,P,history), (1111,S,entertainment), (1111,L,history),(1111,L cosmetics), (2222,L,sports), (2222,S,sports), (3333, S, photography), (3333, L, art), (3333,P,art)

(note the omitted white space in topic names) and the correct output (on the standard output) would be

(1111,history,50), (1111,entertainment,40), (1111,history,20),(1111,cosmetics,20), (2222,sports,20), (2222,sports,40), (3333, photography,40), (3333, art,20), (3333,art,50)

Note that the Mapper program should output each tuple (userID, topic, score) as soon as it processes the corresponding (userID, action, topic) tuple. Also, each tuple should be output on a separate line (although shown in a condensed form above.)

Reducer program gets the tuples of the form (userID, topic, score) from the standard input and generates tuples of the form (userID, topic, total score) on the standard output as follows:

(1111,history,70)
(1111,entertainment,40)
(1111,cosmetics,20)
(2222,sports,60)
(3333,photography,40)
(3333,art,70)

Like Mapper's input, tuples are sorted according to userID field but tuples that belong to the same userID may not be sorted according to the topic. Note that the Reducer outputs the tuple as soon as it realizes that there won't be any more tuples that belong to the same user. You can assume that the Reducer's input has a single tuple 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 that contains tuples in the form (userID, action, topic) and generates tuples of the form (userID, topic, total score). Each tuple is output on a separate line. You should use fork, exec, pipe, and dup2 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 file and preferably a Makefile on CANVAS.