EEL 5733/4732 Advanced Systems Programming
Preliminary Assignment

This is not an assignment. However, if you implement the two programs, Mapper and Reducer, as described in this document, you will be able to reuse your implementation for Assignment 1 and focus on implementing more advanced topics.

Mapper program gets the input from the standard input that contains tuples in the form (userID, action, topic). userID is a 4-digit identification for the users of some 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 maximum 15 characters. The mapper program processes these tuples to generate a weighted profile in the form of (userID, topic, score) based on a set of scoring 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 the 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)

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 (important for better performance!). Also, each tuple should be output on a separate line as shown 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. Assuming that it gets the output of Mapper as shown above as the input, it generates the output by adding up the scores for each topic as follows:

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

Like Mapper's input, the tuples are sorted according to the userID field but tuples that belong to the same userID may not be sorted according to the topic. **Note that the Reducer should output 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.