Scala
COMP 4210
Spring 2021

Project 3

Write a Scala program that solves the following problem.

A company has been hired to collect data that will be used to evaluate a new transportation system. However, their work has to be performed under complete secrecy due to concerns about potential public response to the results of the study.

In order to maintain security, separate teams have been sent out to take measurements along the length of the system. The measurements are also encoded so that what they represent is hidden; the data in each column is disguised to appear to be something unrelated to the actual measurement. Each team is given three things to measure, and to keep their job simple they only record every time one of the measurements changes. For example here is a sample of one of the teams reports:

Start	End	Color	Diameter	Location
1	8	Green	40	San Antonio
8	12	Black	28	Billings
12	15	Orange	40	Norfolk
15	24	Orange	28	Houston
24	25	Red	40	Daytona
25	27	Orange	40	New Orleans

This sample data indicates that from marker 1 to 8 the color, diameter and location values remained constant, but then from marker 8 to 12 they all took on different values. Here is a sample of the report from another team:

Start	End	Height	Verified	Material
1	2	12	FALSE	Concrete
2	5	6	TRUE	Copper
5	6	27	TRUE	Plastic
6	13	48	FALSE	Plastic
13	23	47	FALSE	Steel

As you can see, each of these teams recorded their data with different start and end points for each segment of the line, based purely on where the data they were responsible for monitoring changed in value.

Your job is to collate the data from both input files in order to create a single table with all the data. In order to do this you will need to compute new start stop values that are global to both reports and populate each row with the correct data from both sources. For eaxmple, based on the above tables, your output would begin as follows:

Start	End	Color	Diameter	Location	Height	Verified	Material
1	2	Green	40	San Antonio	12	FALSE	Concrete
2	5	Green	40	San Antonio	6	TRUE	Copper
5	6	Green	40	San Antonio	27	TRUE	Plastic
6	8	Green	40	San Antonio	48	FALSE	Plastic
8	12	Black	28	Billings	48	FALSE	Plastic
12	13	Orange	40	Norfolk	48	FALSE	Plastic
13	15	Orange	40	Norfolk	47	FALSE	Steel

Specifications

Your program should accept the input filenames as command line arguments (not by prompting the user to type them). Each file will be a csv file with 5 columns and will contain a header row that labels each column. Each file will have no more than 5000 rows of data. You should generate a single output file that is also in csv format.

You are required to solve the problem for 2 input files. However, extra credit will be given if your solution accepts an arbitrary number of input files and processes all of them together successfully.

Your goals for this project should be prioritized as follows:

1. A working solution that solves the two-input file version of the problem.

2. Use Scala language features as appropriate to the problem.

3. Formatting: your code should be readable, have clarity, and be well-documented. It should also have your name at the top in comments.

You may feel free to use Scala features we haven't yet covered. However, all code you submit should be your own. Do not look at or in any other way use other students' code.

DUE: Friday, March 19 by 11:59 pm