Description

Xiaoyu Zheng

1. Directory structure

In Solution directories, there are 2 folders, which contain the python one and the scala one separately. In the python folder, there is a python file, while in the scala folder, there are scala source code and the executable jar. In OutputFiles folder, there exist all the output txt files required in all three problems folders. As the files created by python codes and scala codes are the same, so I just provide one version for each output txt file.

2. Execution rules

1) Put my **executable files** (like Xiaoyu_Zheng_SON.py, or Xiaoyu_Zheng_SON.jar) directly **in the root directory of Spark**.
2) Put all the input **csv files** in the **root directory of Spark**, and **change their names**.
   Change the name of the **ratings.csv** file in **ml-latest-small** folder to **MovieLens.Small.csv**
   Change the name of the **ratings.csv** file in **ml-20m** folder to **MovieLens.Big.csv**
3) To utilize my work, please **get into the root directory of spark** with terminal or in the command line.
4) Then to execute my work, you need the following commands. Note: These commands will overwrite the files which have the same name as my output files in the root directory of Spark.

   Python:
   bin/spark-submit Xiaoyu_Zheng_SON.py <case number> <csv file name> <support>
   For example:
   bin/spark-submit Xiaoyu_Zheng_SON.py 1 Small1.csv 4

   Scala:
   bin/spark-submit --class FrequentItemsetsSON Xiaoyu_Zheng_SON.jar <case number> <csv file name> <support>
   For example:
   bin/spark-submit --class FrequentItemsetsSON Xiaoyu_Zheng_SON.jar 1 Small1.csv 4

5) The output txt files will also be in the root directory of Spark.