953212 DB SYS & DB SYS DESIGN 68/1

(Lab1) Dealing with traditional file with Java (5 points)

FirstName _Samakom_ Lastname _Prompanya_ student ID _672115047_

In this lab, you are going to apply your programming and problem-solving skills to generate small report from one given file, namely 2024 QS World University Ranking 1.1 (For qs.com).csv

Information about data are describes as follows:

Part 1) Data Set Information and Attribute Information:

1) (1 point) What are the meta data for the given data? please list all

   : List all the metadata – e.g., attribute names, descriptions, units if available, etc.

   2024 Rank, 2023 Rank, Institution Name, Country Code, Country, Size, Focus, Res., Age, Status, Academic Reputation Score, Academic Reputation Rank, Employer Reputation Score, Employer Reputation Rank, Faculty Student Score, Faculty Rank, Citations per Faculty Score, Citations per Rank, International Faculty Score, International Faculty Rank, International Student Score, International Student Rank, International Research Network Score, International Research Rank, Employment Outcomes Score, Employment Outcomes Rank, Sustainability Score, Sustainability Rank and Overall Score.

2) (1 point) List all the data type of each attribute.

   : 2024 Rank(Integer), 2023 Rank(Integer), Institution Name(String), Country Code(String), Country(String), Size(String (Category)), Focus(String (Category)), Res.(String (Category)), Age (Integer), Status(String (Category)), Academic Reputation Score(Float), Academic Reputation Rank(Integer), Employer Reputation Score(Float), Employer Reputation Rank(Integer), Faculty Student Score(Float), Faculty Rank(Integer), Citations per Faculty Score(Float), Citations per Rank(Integer), International Faculty Score(Float), International Faculty Rank(Integer), International Student Score(Float), International Student Rank(Integer), International Research Network Score(Float), International Research Rank(Integer), Employment Outcomes Score(Float), Employment Outcomes Rank(Integer), Sustainability Score(Float), Sustainability Rank(Integer) and Overall Score(Float).

Part 2) Basic descriptive with Java

1) (1 point) Write a Java program to open, read, and close the given CSV file.
2) (1 point) Insert Chiang Mai University to the record in year 2024 and write to file as QS_World University Ranking 2024.csv

More info can be found here:

https://www.topuniversities.com/universities/chiang-mai-university

3) **[Extra credit]** What problem have you found when you are dealing with the tradition database files? Explain in detail? (e.g., domain integrity, NULL values, etc.) (2 points)

: > Missing or NULL values :

Some columns have no data. For example, some universities don't provide scores in certain areas, which makes calculations and analysis difficult.

 > Invalid or inconsistent values (domain issues)

Some data are outside the expected range. For example, the "SIZE" column should only have values like S, M, L, or XL, but sometimes it contains unknown values or blanks.

 > Inconsistent data formats :

Some columns use different formats, such as "US" and "United States" in the country column. This makes grouping and filtering data more difficult.

 > No built-in data validation :

Files like CSV don't have constraints like real databases (e.g., primary keys, not null, unique). So, it's easy to have duplicate rows or incorrect data.

Submission

Submit your Java source code Main.java, UniData.java to Mango under Lab1.

You can write Part 1/Part 3 answer in your Main.java using comment.

Late submission -2