

Practical Assignment for Junior Java Developer

Summary

Develop a command line tool in **Java** that is able to merge data from two CSV files. The tool should be able to identify the common information in the files and based on that, generate 3 new CSV files. The first one will contain the data that is contained in **both** input files, the second one will contain the data that can be found only in the **first** input file and the last one will contain the data that can be found only in the **second** file.

Input

The program should take two paths to the input CSV files. These two paths shall be provided as arguments at the command line.

Output

The program will generate 3 files in the directory from where the program was run:

- “common.csv”: A CSV file containing the rows that could be found in both the first and second CSV files.
- “diff1.csv”: A CSV file containing the rows that could be found in the first CSV, but not in the second.
- “diff2.csv”: A CSV file containing the rows that could be found in the second CSV, but not in the first.

Specification

The first row in a CSV file is always presumed to be the “header row”. Thus, every value on the first row is the name of a column. The other rows starting from the second are presumed to be data. Each data row will have the same number of columns as the header row.

When comparing two rows from two CSV files:

- Consider only the values associated with common columns in both files.
- Two rows are equal if they share the exact same values on all of the common columns.

The common CSV file should merge the equal rows. Therefore, the common CSV file may have more columns than any of the input files.

Example

The following is an example input. The left one is the content of the first CSV file and the right one is the content of the second CSV file.

Title,Author,Year	Author,Pages,Cover,Title
The Great Gatsby,F. Scott Fitzgerald,1925	Charlotte Brontë,532,Paperback,Jane Eyre
Jane Eyre,Charlotte Brontë,1847	Daniel Defoe,320,Hardcover,Robison Crusoe
Frankenstein,Mary Shelley,1818	Jane Austen,279,Paperback,Pride and Prejudice
Robison Crusoe,Daniel Defoe,1719	

The common columns here are only **Title and Author**. The following are the contents of the 3 output files (diff1.csv, common.csv and diff2.csv):

Title,Author,Year	Title,Author,Year,Pages,Cover	Author,Pages,Cover,Title
The Great Gatsby, F. Scott Fitzgerald,1925	Jane Eyre,Charlotte Brontë, 1847,532,Paperback	Jane Austen,279,Paperback, Pride and Prejudice
Frankenstein, Mary Shelley,1818	Robison Crusoe,Daniel Defoe, 1719,320,Hardcover	

Note that 2 records were exclusive to the first file and 1 record was exclusive to the third file. Finally, 2 records were common, so we merged them into a single record. Mind that the example wrapped lines - the program shouldn't care for any line wrapping in the process. Also, the example uses colors and bold styling - the program shouldn't care for any highlighting.

Reference

- You can consult the available online documentation for Java and 3rd party libraries you are going to use.
- You are **not** allowed to use: AI assistants, online or offline guidance

Copyright

- The goal of the information here is to provide a context for an interview project that is to be developed by an Atek Software candidate for the Junior Java Developer position.
- This document is part of the Atek Software intellectual property. It is forbidden to be altered, copied or reproduced in any way, without author's permission. This document **shouldn't** be deliberately shared to other persons.
- Atek Software is allowed to use or share the implementation as long as it doesn't correlate it back to the candidate. In the same way, the candidate is allowed to use or share the implementation as long as it doesn't correlate it back to the Atek Software company.