

Web Table Extraction Task

In the “data/” directory you will find four HTML files. Your task is to extract all tables from these files and save their content in CSV files. Create one CSV file per table. As a simple heuristic which tables should be used, apply the following rules:

- We are not interested in tables containing other tables
- Tables must contain at least 2 rows and 2 columns

The CSV files should be as clean as possible (i.e., should not contain HTML artefacts). You should also think about how to handle spanning cells in the tables (in the CSV, each line must have the same number of values).

This task must be completed in Java, you’re free to use any library that you find useful. Please hand in your source code (everything that’s needed to compile and run your solution), the extracted CSV files and a short description of your approach.