

The Ninth International Conference on Advances in Databases, Knowledge, and Data Applications, Mai 21 - 25, 2017 - Barcelona, Spain

**Tutorial: Data Manipulation and Data Transformation using the Shell
Andreas Schmidt, Steffen Scholz**

Exercise III

Tasks:

1. Create a working copy of your file `city.csv` (for security reasons)
2. Exchange in the file all occurrences of the Province "Amazonas" in Peru (Code PE) with "Province of Amazonas" using `sed` (inplace).
3. Look for entries with the String "ce of Amazonas" - it should be only 1 !
4. Make the same operations using `awk`.
5. Print all cities which have no population given.
6. Print the line numbers of the cities in Great Britain (Code: GB) using `awk`.
7. Delete the records 5-12 and 31-34 from `city.csv` and store the result in `city.2.csv` using `sed`.
8. Combine the used commands from the last two tasks and write a bash-script (sequence of commands), which delete all british cities from the file `city.csv` (Hint: generate with `awk` the commands for `sed` to delete the corresponding lines)
9. Count the datasets (lines) in `city.csv` - it should be 2880
10. If you take a look at the files, downloaded from the Gutenberg Project, you can identify some boilerplate text at the begin and the end of the book. Which are the lines, who separate the literary text from the boilerplate text?
11. Write a command, which removes the boilerplate text (Hint: use `sed`, `head`, `tail`)