

Linux Left Outer Join 2er CSV Dateien

Das Ziel hier ist es, 2 CSV-Files, deren Daten durch ';' getrennt sind zu einer gemeinsamen Tabelle zu fusionieren.

Dabei sollen alle Zeilen der *bom2* auftauchen, egal, ob *parts* ein Gegenstück enthält oder nicht.

Ausgegeben werden alle Spalten.

```
#!/bin/bash

# ACHTUNG: dieses problem lässt sich leider nicht lösen.
# es gibt 1 Problem:
# 2. join/sort können keine " als escape-zeichen verstehen.

# macht einen left outer joint zwischen bom2.csv und parts.csv

# die bom wird vom footer befreit
# (d wie data only)
head -n -15 bom2.csv > bom2d.csv

# dann müssen wir die tabellen sortieren
# ganz wichtig hier: das stop-feld
#
https://stackoverflow.com/questions/25431673/linux-join-utility-complains-about-input-file-not-being-sorted#25432076
# also key=7,7 statt key=7 (!!!)
#
# zusätzlich überspringen wir die 1. Zeile (Header)
# siehe
https://stackoverflow.com/questions/14562423/is-there-a-way-to-ignore-header-lines-in-a-unix-sort
#
# LOCALE sorg dafür, dass die sortierung bei sort und join einheitlich ist.
#
(LOCALE=C head -n 1 bom2d.csv && tail -n +2 bom2d.csv | sort --field-separator=';' --key=7,7) > bom2s.csv
(LOCALE=C head -n 1 parts.csv && tail -n +2 parts.csv | sort --field-separator=';' --key=1,1) > partss.csv

# nun kommt der join

# -t \;
# gitb den Strichpunkt als Trennzeichen der Ein- und Ausgabedateien an.
```

```
# -1 7
# Mache den Join mit der 7. Spalte von Datei 1

# -2 1
# mache den Join mit der 1. Spalte von Datei 2

# -a 1
# gib alle Zeilen von Datei 1 aus, auch wenn kein Treffer gefunden wurde

# -e MISSING
# Fülle alle Felder, die Fehlen, mit MISSING auf

# -o auto
# Ausgabeformat, alle Spalten von 1. und 2. Datei

# > out.csv
# Speichere das Ergebnis in out.csv

LOCALE=C join --header -t \; -1 7 -2 1 -a 1 -o auto bom2s.csv partss.csv >
bom3.csv

# Aufräumen
rm bom2d.csv
rm bom2s.csv
rm partss.csv
```

From:

<http://www.zeilhofer.co.at/wiki/> - **Verschiedenste Artikel von Karl Zeilhofer**

Permanent link:

http://www.zeilhofer.co.at/wiki/doku.php?id=linux_left_outer_joinLast update: **2017/08/03 20:41**