

Umgestalten einer Tarifliste

Situation: Vor ein paar Jahren kam ein Nachbar mit einer Tarifliste für Autogarage zu mir. Er erklärte mir, dass jede Arbeit in der Garage nach einem Tarifbuch verrechnet werden. Jede Arbeit, zum Beispiel der Ersatz einer Karosserieseitenwand hat eine bestimmte Nummer und gehört zu einer bestimmten Gruppe von Arbeiten.

Die Mechaniker in der Werkstatt notieren die Arbeit und das Büro muss danach die Notizen mit dem Tarifbuch abgleichen und erfassen.

Es wäre wesentlich einfacher, wenn der Mechaniker die Arbeit auf der Basis des Tarifbuches erfassen würde. Dazu sei Liste aller möglichen Arbeiten zu gross und zu unübersichtlich. Zu viele Spalten, zu viele Sprachen. Die Mechaniker finden sich in der Liste nicht zurecht.

Die Liste muss vereinfacht werden:

- nur noch die deutschen Texte
- nur noch die Hauptgruppen
- nur noch zwei bestimmte Spalten

Das Tarifbuch wird immer mal wieder überarbeitet. Deshalb soll die Vereinfachung mit einem Script automatisiert werden.

Sie lernen:

- Systematisches Vorgehen beim Erarbeiten eines Scripts
- Die Anwendung von grep, cut, (sort, wc)
- Umgang mit grossen Datenmengen

Auszug aus der Tarifliste (Export.spl)

So sehen die Daten aus bevor das Script durchgelaufen ist

```
2 0 'A001500'      'Teilersatz Karosserieseitenwand vorn - Rechte Seite'
2 0 'A001500001'   'Schutzzierleiste an Seitenwand vorn anbringen'
2 0 'A001500005'   'B-Säule oberhalb ersetzen'
2 0 'A001600'      'Teilersatz Karosserieseitenwand vorn - Linke Seite'
2 0 'A001600001'   'Schutzzierleiste an Seitenwand vorn'
...
7 0 'A001500'      'Sostituzione parziale della fiancata anteriore - Lato destro'
7 0 'A001500001'   'Applicazione della modanatura protettiva sulla fiancata
                    anteriore'
7 0 'A001500005'   'Sostituzione del montante B in alto'
```

Erwartetes Resultat

So sollen die Daten aussehen, nachdem das Script sie verarbeitet hat.

```
'A001500'      'Teilersatz Karosserieseitenwand vorn - Rechte Seite'
'A001600'      'Teilersatz Karosserieseitenwand vorn - Linke Seite'
'A201598'      'Bodenblech vorn abdichten'
'A203098'      'Bodenblech mitten abdichten'
```

Script Aufruf

So soll das Script aufgerufen werden können:

```
lp5prutschmann@SYSW4220 ~
$ ./convertData.sh Export.spl output.txt
```

mit und ohne Parameter

```
$ ./convertData.sh
Dieses Script filtert aus einer speziellen Liste
die deutschsprachigen Hauptarbeitsbezeichnungen heraus.

Sie muessen beim Aufruf des Programm ./convertData.sh den Namen der Input-,
und der Output-Datei mitgeben.
./convertData.sh [input] [output]
```

Ein guter Weg für die Lösung ist ein Schritt für Schritt Vorgehen:

Schritt 1 „Prüfen von Input und Output Parameter“

- Schreiben Sie ein Script *convertData01_Parameter.sh*, dem zwei Parameter beim Start übergeben werden.
- Das Script prüft die beiden Parameter, auf dass die beiden nicht leer sind.

Ausgaben:

```
peter.rutschmann@N001 ~  
$ ./convertData01_Parameter.sh Export.sp1 output1.txt  
input:  Export.sp1  output:  output1.txt
```

Ausgabe im Fehlerfall:

```
peter.rutschmann@N001 ~  
$ ./convertData01_Parameter.sh Export.sp1  
Dieses Script filtert aus einer speziellen Liste  
die deutschsprachigen Hauptarbeitsbezeichnungen heraus.  
  
Sie muessen beim Aufruf des Programm ./convertData01_Parameter.sh den Namen der Input-,  
und der Output-Datei mitgeben.  
./convertData01_Parameter.sh [input] [output]
```

Schritt 2 „Alle deutschsprachigen Texte herausfiltern“

- Schreiben Sie ein Script *convertData02_alleDeutschen.sh* auf der Basis des Scripts *convertData01_Parameter.sh*.
- Mittels **grep** filtert das Script alle deutschsprachigen Texte aus der [input] Datei und gibt das Resultat an in die Dateien [output] aus.

Hilfe:

Betrachtet man die Tarifliste *Export.sp1*, so sieht man:

- Alle Zeilen mit deutschen Texten beginnen mit einer 2, dann folgt ein Tabulator und dann eine beliebige Zahl dann kommt der Rest der Zeile.
- | | | | |
|---|---|--------------|------------------------|
| 2 | 0 | 'A315200001' | 'C-Säule ersetzen' |
| 4 | 1 | 'A315200001' | 'Montant C - remplace' |
- Also muss man alle Zeilen, die dieser Regel entsprechen herausfiltern
„Zeile beginnt mit 2“

```
peter.rutschmann@N001 ~  
$ ./convertData02_alleDeutschen.sh Export.sp1 output2.txt
```

Leider gibt es auch Zeilen mit einer 2 am Anfang, die nicht deutsch sind.

2	0	'C001000'	'Windschutzscheibe aus- und einbauen oder ersetzen'
2	26	'C001000'	'Windscreen - Remove and Reinstall or Replace'

Somit kommen wir zum Schritt 3

Schritt 3 „Alle Zeilen mit einer „26“ rausfiltern

- Schreiben Sie ein Script `convertData03_alleOhne26.sh` auf der Basis des Scripts `convertData02_alleDeutschen.sh`.
- Zusätzlich sollen die „falschen“ Zeilen rausgefiltert werden. (Mittels `grep`)

Hilfe:

Betrachtet man das Zwischenresultat so sieht man:

- Alle falschen Zeilen beginnen mit einer 2, dann folgt ein Tabulator und dann eine 26

```
2      26      'C001000'      'Windscreen - Remove and Reinstall or Replace'
```

- 1) Man filtert alle Zeilen, die mit einer 2 beginnen, mittels `grep`.
- 2) Dann leitet man den output dieses Filters mittels Pipe an einen zweiten Filter weiter.
- 3) Der zweite Filter filtert alle Zeilen, die NICHT der zweiten Regel entsprechen.
grep filtere alles, was **NICHT** der Regel „2, dann folgt ein Tabulator und dann eine 26 entspricht“

```
peter.rutschmann@N001 ~  
$ ./convertData03_alleOhne26.sh Export.sp1 output3.txt
```

Schritt 4 „Spalte 1 und 2 sollen verschwinden“

- Schreiben Sie ein Script `convertData04_SpaltenWeg.sh` auf der Basis des Scripts `convertData03_alleOhne26.sh`.
- Zusätzlich sollen die ersten zwei Spalten entfernt werden. (Mittels `cut`)

```
2      0      'A001500'      'Teilersatz Karosserieseitenwand vorn - Rechte Seite'  
  
'A001500'      'Teilersatz Karosserieseitenwand vorn - Rechte Seite'
```

Hilfe:

- „Schneide“ die Spalten mittels `cut`
- Verwende wiederum eine Pipe

```
peter.rutschmann@N001 ~  
$ ./convertData04_SpaltenWeg.sh output2.txt output4.txt
```

Schritt 5 „Nur die Hauptgruppen ausgeben“

- Schreiben Sie ein Script `convertData05_NurHauptgruppen.sh` auf der Basis des Scripts `convertData04_SpaltenWeg.sh`.
- Es sollen nur noch die Hauptgruppen aufgeführt werden. (Mittels `grep`)

```
'A001500'      'Teilersatz Karosserieseitenwand vorn - Rechte Seite'
```

Hauptgruppen beginnen mit A..Z dann folgt eine 6 stellige Zahl

Hilfe:

- Filtere alle Zeilen, die NICHT „Mit A..Z beginnen und dann 7 Zifferen haben“ mittels **grep**

```
peter.rutschmann@N001 ~  
$ ./convertData05_NurHauptgruppen.sh Export.sp1 output5.txt
```