

– Lösung zur Praktikumsaufgabe 4 –

Thema: *Reguläre Ausdrücke*

1. Es ist durchaus zulässig, sich zunächst mit den im vorigen Praktikum beschriebenen Mitteln eine Textdatei anzulegen, die nur die Bibelverse (einen pro Zeile enthält). Dies kann z. B. mittels

```
~> cut -d ' ' -f 3- bibel.txt > verse.txt
```

erfolgen. Eine solche Datei namens `verse.txt` wird für die Lösung vorausgesetzt. Das folgende Shellskript bietet für jede Teilaufgabe zumindest eine Lösung:

```
#!/bin/bash

# a)
echo Alle dreibuchstabigen Worte:
grep -o "\<[:alpha:]\{3\}>" verse.txt | sort | uniq
echo Alle vierbuchstabigen Worte:
grep -o "\<[:alpha:]\{4\}>" verse.txt | sort | uniq
# b)
echo -n Anzahl verschiedener Versanfänge:
grep -o "^[:alpha:]+\>" verse.txt | sort | uniq | wc -l
# c)
echo -n Anzahl verschiedener Substantive:
grep -o "\<[:upper:][:alpha:]*\>" verse.txt | sort | uniq | wc -l
# d)
echo Alle Verse, die \(\mindestens\) zwei Löwen enthalten:
grep "Löwe.*Löwe" verse.txt
echo Alle Verse mit drei und mehr Löwen:
grep "Löwe.*Löwe.*Löwe" verse.txt
# e)
echo Das längste Wort der Bibel:
grep -o "\<[:alpha:]\+>" verse.txt | awk '{print length " " $1}' |
  sort -n -r | uniq | head -n 1 | cut -d ' ' -f 2
echo Die 40 längsten Worte der Bibel:
grep -o "\<[:alpha:]\+>" verse.txt | awk '{print length " " $1}' |
  sort -n -r | uniq | head -n 40 | cut -d ' ' -f 2
echo Die längsten Nichtzahlworte der Bibel:
grep -o "\<[:alpha:]\+>" bibel.txt.utf8 | awk '{print length " " $1}' |
  grep -v "\(eins\)\|\(zwei\)\|\(drei\)\|\(vier\)\|\(fünf\)\|\(sechs\)\|\(sieben\)\|\(acht\)\|\(neun\)\|\(zehn\)\|\(hundert\)\|\(tausend\)" | sort -n -r | uniq | less
# f)
echo Alle Worte, die mit 'g' beginnen und 'n' enden:
grep -o "\<[Gg][:alpha:]*n\>" verse.txt | sort | uniq
echo Alle Worte, die mit 'a' beginnen und mit 'ing' oder 'ung' enden:
grep -E -o "\<[Aa][:alpha:]*[ui]ng\>" verse.txt | sort | uniq
Oder auch:
grep -o "\<[Aa][:alpha:]*\(\(ing\)\|\(ung\)\)\>" verse.txt | sort | uniq
```

Anmerkungen:

- Mit `head -n x <datei>` gibt man die ersten `x` Zeilen der Datei `<datei>` aus.
- Das längste Wort der Bibel, das keine Zahl ist, ist *Hohenpriestergeschlecht* auf Platz 34.
- Die letzte Zeile ist ein Beispiel für einen so genannten erweiterten regulären Ausdruck (er kann aber ohne weiteres durch intensives Einstreuen von Backslashes in einen *basic regexp* transformiert werden)!

2.

```
#!/bin/bash
while true
do
    echo -e `date +%T`
    sleep 1
    clear
done
```

Es tauchte im Praktikum die Frage auf, ob man auch ohne komplettes Löschen des Terminals auskommen kann. Die Nutzung des Steuerzeichens `\r` (*Carriage Return*; Wagenrücklauf) in Verbindung mit `echo` wäre eine Alternative:

```
echo -e `date +%T` \r
```

3. `uniq -c` gibt die Häufigkeit der gefundenen Zeilen aus.

```
#!/bin/bash
cut -f1 -d' ' ~/.bash_history | sort | uniq -c | sort -n -r | less
```