

Programmiertechniken in der Computerlinguistik I

Universität Zürich, September 2019

Martin Volk, Simon Clematide

Aufgabe 0

Anwenden von Unix-Kommandos

Probiere die folgenden Unix-Kommandos:

Befehl	Bedeutung	Erklärung
ls -l	listing	Auflisten der Dateien in einem Verzeichnis
pwd	print working directory	Zeige den aktuellen Pfad und das aktuelle Verzeichnis
cd	change directory	Wechsle in ein anderes Verzeichnis
mkdir	make directory	Erzeuge ein Unterverzeichnis im aktuellen Verzeichnis
wc	word count	Zeige die Grösse einer Datei: Anzahl Zeilen, Wörter und Zeichen
mv	move	Verschiebe eine Datei aus einem Verzeichnis in ein anderes Verzeichnis; Ändere den Namen einer Datei
rm	remove	Lösche eine Datei. Vorsicht! Die Datei ist dann unwiederbringlich verloren.
cp	copy	Kopiere eine Datei.
touch		Weise einer Datei das aktuelle Datum zu.
chmod	change modalities	Verändere die Zugriffsrechte auf eine Datei für Benutzer, Gruppe, Andere (user, group, others).
sort		Sortiere die Zeilen einer Datei.
uniq -c		Fasse nebeneinander liegende identische Zeilen einer Datei zusammen und zähle die Häufigkeit jeder Zeile.
grep		Finde einen Suchbegriff in einer Zeile
tr	transliterate	Ersetze einen String mit einem anderen String

1. Mach dich mit dem Inhalt der im Materialordner bereitgestellten Datei uebung00.txt bekannt.
Probiere die obigen Kommandos mit deiner Testdatei.
2. In OLAT liegen die SAC-Bände «Die Alpen» von 1930 bis 1939 im 3-Spalten-Format (Wortform \t PoS-Tag \t Lemma) für dich bereit. Lade diese auf deinen Computer.
 - 2.1. Sortiere die Zeilen im Band von 1935 alphabetisch. Berechne die Anzahl der unterschiedlichen Wörter (~Zeilen) in der Datei. Welches sind die 10 häufigsten Wörter?
 - 2.2. Berechne die Anzahl der unterschiedlichen Wörter über allen 10 Bänden zusammen. Welches sind die 10 häufigsten Wörter?

2.3. Durchsuche die 10 Bände mit Hilfe von grep. Beispielfragestellungen:

- Wie oft kommt das Wort «Freund» (=die Wortform) in den 10 Bänden vor? Wie oft kommt das Wort «Kamerad» vor?
- Wie oft kommt das Lemma «Freund» in den 10 Bänden vor? Wie oft kommt das Lemma «Kamerad» vor?
- In welchem Band kommt das Lemma «Freund» am häufigsten vor?
- Berechne die Anzahl der unterschiedlichen Wörter in der Datei. Welches sind die 10 häufigsten Wörter?
- Wie viele Verben kommen in den 10 Bänden vor?
-

Ziel:

- Kennenlernen der Bedienung eines Eingabefensters (Terminal)
- Kennenlernen der wichtigsten UNIX-Kommandos
- Kennenlernen des Zusammenfügens von Kommandos (Pipe |) und des Umlenkens der Ausgabe in eine Datei(>)

Einzureichen sind:

- Diese Aufgabe ist nur zu Deinem eigenen Einstieg in den Kurs. Du musst nichts einreichen.