

Übung 11

Extraktion von Daten aus XML-Dateien

Themen: reguläre Ausdrücke

Extrahieren Sie aus XML-Dateien mit annotierten Texten die Daten und speichern Sie Token, Wortart mit morphosyntaktischen Merkmalen und Lemmata in einer dreispaltigen TSV-Datei mit Tabulatoren als Trennzeichen wie in folgendem Beispiel:

```
ist      VAFIN.Ind.Pres.Sg.3      sein
```

Einen Link zu den Daten finden Sie auf der Kursseite. Verwenden Sie reguläre Ausdrücke zur Extraktion der Daten.