# Übung 2

In dieser Übung schreiben Sie Code zur Zählung von Häufigkeiten, zur Schätzung von Wahrscheinlichkeiten und zur Berechnung von Entropiewerten.

**Aufgabe 1:**

Laden Sie die Daten in `https://www.cis.lmu.de/~schmid/lehre/data/tiger.txt.gz` herunter und dekomprimieren Sie die Datei.

Schreiben Sie dann eine Funktion `read_corpus(filename)`, welche einen Dateinamen als Argument erhält und das darin enthaltene getaggte Korpus einliest. Die Datei enthält eine Spalte mit Wörtern und eine Spalte mit Tags. Nach jedem Satz folgt eine Leerzeile.

Die Funktion extrahiert die Häufigkeiten aller Tags und aller Wort-Tag-Paare und gibt die Häufigkeiten in zwei Dictionaries zurück.

**Aufgabe 2:**

Schreiben Sie eine Funktion `estimate_probs(freq)`, welche ein Dictionary `freq` mit Häufigkeiten als Argument erhält und ein Dictionary mit Wahrscheinlichkeiten zurückliefert.

**Aufgabe 3:**

Schreiben Sie eine Funktion `estimate_cond_probs(freq)`, welche ein Dictionary `freq` mit den Häufigkeiten von Wort-Tag-Paaren als Argument erhält und ein Dictionary mit den bedingten Wahrscheinlichkeiten der Tags gegeben die Wörter zurückgibt.

**Aufgabe 4:**

Schreiben Sie eine Funktion `entropy(prob)`, welche ein Dictionary `prob` mit Tags als Keys und ihren Wahrscheinlichkeiten als Values bekommt, und die Entropie der Wahrscheinlichkeitsverteilung berechnet.

Die berechnete Entropie wird zurückgegeben.