

WiSe 2016/2017

# Texttechnologie 1

## 1. Übungsblatt

28. Oktober 2015

Abgabefrist: 11. November 2015, 16 Uhr

### 1 Goethe vs. Schiller

Im [Projekt Gutenberg](#) sind natürlich auch die beiden deutschen Nationaldichter Goethe und Schiller vertreten. In der Übung sollen Sie beide Autoren exemplarisch quantitativ miteinander vergleichen, und zwar anhand zweier berühmter Beispieltex-te, nämlich den [Wahlverwandschaften](#) und den [Räubern](#) (beide Texte sind in aufgeräumter Form als Textdateien im [OLAT](#) hinterlegt).

Schreiben Sie ein Programm, das die Texte tokenbasiert einliest (Achtung: Umlaute!). Erstellen Sie für jeden der beiden Text eine alphabetische Liste seiner Types (Wortformen), unter Ausschluss von Satzzeichen u. ä. Welches sind die 20 häufigsten Wortformen? Welches sind die 20 häufigsten großgeschriebenen Wortformen? Finden Sie Fälle, die Sie für problematisch halten? (5 Pkte)

### 2 Type-Token

Erstellen Sie für beide Texte die *Type-Token-Ratio* (TTR) und die *Moving-Average Type-Token-Ratio* (MATTR; Fenstergröße 300 Token). Unterscheiden sich die Ergebnisse? Wie erklären Sie den Unterschied zwischen TTR und MATTR? (5 Pkte)

### 3 Über den Tellerrand

Schauen Sie, ob es ein externes Tool oder ein Paket für Ihre Programmiersprache gibt, dass die Routinen des Texteinlesen und der Berechnung von TTR/MATTR aus den Abschnitten 1 und 2 bereit stellt. Replizieren Sie die Aufgabe mit diesem Tool oder Programm. (5 Pkte)

**Abgabe und Punkte.** Geben Sie Ihr Programm (kommentierter Quellcode, Anleitung zur Benutzung und zur Erzeugung der Ausgaben, **und** lauffähiges Programm (z. B. `jar`)) sowie die Ergebnisausgaben und eine knappe Antwort auf oben gestellte Fragen fristgerecht als komprimiertes Archiv im [OLAT](#) ab (max. 15 Pkte.).