

Übung zur Vorlesung “Einführung in die Computerlinguistik und Sprachtechnologie”

Wintersemester 2018/2019, Prof. Dr. Udo Hahn, Luise Modersohn

Übungsblatt 10 vom 31.01.2019

Abgabe bis 06.02.2019, 23.59 Uhr; per Email (PDF-Format) an

luise.modersohn@uni-jena.de

Aufgabe 1 kMeans Clustering

2

Stellen Sie knapp das Vorgehen beim k-Means-Clustering dar. Was sind die wesentlichen Schritte? Wie häufig werden sie durchgeführt?

Aufgabe 2 Grammatiken

Schreiben Sie eine Grammatik, die arithmetische Ausdrücke folgenden Typs erzeugen kann:

- $2 + 3$
- $2 + 3/4$
- $((2 + 3) + 7)/33$

Die Ausdrücke enthalten also folgende Elemente: ganze Zahlen (≥ 0), die Operatoren der vier Grundrechenarten sowie (jeweils paarweise) Klammern. Begründen Sie welchen Typs diese Grammatik mindestens sein muss (und verwenden Sie diesen auch).

Aufgabe 3 Parsing

6

Gegeben sei eine Typ-2-Grammatik mit folgenden Regeln:

$S \rightarrow NP VP$

$S \rightarrow VP$

$NP \rightarrow n$

$NP \rightarrow det n$

$VP \rightarrow v NP$

$n \rightarrow \text{Parse} \mid \text{Satz}$

$v \rightarrow \text{Parse}$

$det \rightarrow \text{den}$

Führen Sie bei dem Satz "Parse den Satz" jeweils einmal Top-Down- (Tiefensuche), Bottom-Up- und Left-Corner-Parsing durch.

Hinweise. Zeigen Sie für jede Regelanwendung den Zwischenstand des Parsing-Prozesses in einem separaten Bild. Wenn Sie sich zwischen mehreren Regeln entscheiden müssen, arbeiten Sie die Kandidaten von oben nach unten ab, so wie sie in der Grammatik stehen. Im Fall von Backtracking, machen Sie explizit deutlich, zu welcher Entscheidung zurückgekehrt wird.