

# Übungen zur Vorlesung Sequenzanalyse

Universität Bielefeld, WS 2013/2014  
Prof. Dr. Jens Stoye · Linda Sundermann

<http://wiki.techfak.uni-bielefeld.de/gi/Teaching/2013winter/SequenzAnalyse>

## Übungsblatt 8 vom 10.12.2013

Abgabe in einer Woche vor Beginn der Vorlesung.

### Aufgabe 1 ( $\Sigma$ -Baum)

(4 Punkte)

Gegeben sei die Menge der Worte  $W = \{cars, sad, salad, arc, card, at, scar, cat\}$ .

1. Zeichne den kleinsten  $\Sigma$ -Baum und den kleinsten kompakten  $\Sigma^+$ -Baum, welche alle Worte aus  $W$  darstellen.
2. Gib jeweils die Menge der Worte  $x \in \Sigma^*$  an, für die  $node(x)$  definiert ist.
3. Welche Menge  $words(T)$  von Worten wird durch die Bäume dargestellt?

### Aufgabe 2 (Suffixbäume)

(4 Punkte)

1. Beschreibe zwei Anwendungen aus der bioinformatischen Praxis, für die man einen Suffixbaum verwenden kann.
2. Gegeben ist der String  $s\$ = abcacbbaca\$$ , wobei  $\$ < a < b < c$ .
  - (a) Zeichne den Suffixbaum für  $s$ , sortiere dabei die Blätter lexikographisch.
  - (b) Beschrifte die Blätter mit dem Start-Index des zugehörigen Suffixes in  $s$ . Die Indizierung beginnt bei 1.
  - (c) Beschrifte jeden Knoten mit der Anzahl der unter ihm liegenden Blätter.

### Aufgabe 3 (WOTD-Algorithmus)

(4 Punkte)

1. Führe den WOTD-Algorithmus schrittweise zur Erstellung des Suffixbaums für den String  $t\$ = BABBA\$$  aus (so wie es im Skript auf Seite 60 gezeigt ist).
2. Wie ergeben sich *worst case* und *average case* Laufzeit und wie lauten sie?
3. Welche Eigenschaften machen den WOTD-Algorithmus in der praktischen Durchführung interessant?