

# Übungen zur Vorlesung Sequenzanalyse II

Universität Bielefeld, SoSe 2012

Dr. Alexander Sczyrba · Nina Luhmann · Linda Sundermann

<http://wiki.techfak.uni-bielefeld.de/gi/Teaching/2012summer/SequenzAnalyse>

## Übungsblatt 3 vom 27.04.2012

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 (Suffixbaum)

(4 Punkte)

1. Zeichne den Suffixbaum von  $s = \text{AACTGACAACGA}\$$ . Sortiere die Blätter lexikographisch, wobei  $\$ < A < C < G < T$ , und beschrifte jedes Blatt mit dem Startindex des zugehörigen Suffixes in  $s$ . Fange bei 0 an zu indizieren.
2. Gib eine Familie von Strings  $s_n^{\max}$  der Länge  $|s_n^{\max}| = n$  an, so dass der kompakte Suffixbaum von  $s_n^{\max}\$$  maximal viele innere Knoten hat. Weiterhin gib eine Familie von Strings  $s_n^{\min}$  der Länge  $|s_n^{\min}| = n$  an, so dass der kompakte Suffixbaum von  $s_n^{\min}\$$  minimal viele innere Knoten hat. Das Alphabet darf dabei beliebig gewählt sein und von  $n$  abhängen.

### Aufgabe 3 (Anzahl innerer Knoten)

(4 Punkte)

Erkläre in eigenen Worten, dass folgendes Lemma gilt:

*Der Suffixbaum eines Strings der Länge  $n$  hat höchstens  $n - 1$  innere Knoten.*