

# Übungen zur Vorlesung Sequenzanalyse

Universität Bielefeld, SS 2019

Dr. Daniel Dörr · Michel T. Henrichs

<https://gi.cebitec.uni-bielefeld.de/teaching/2019summer/sa>

**Übungsblatt 9 vom 3.6.2019**

**Abgabe am 11.6.2019 bis 10:00 Uhr**

## Aufgabe 1 (Maximale Repeats)

(6 Punkte)

1. Finde alle maximalen Repeats in  $s = \text{AGTGCAATGTGCAT}$  unter Verwendung des im Skript geschilderten Algorithmus (Abschnitt 7.7.3).
2. Argumentiere unter Berücksichtigung des Suffixbaums, warum folgende Aussage korrekt ist:  
*In jedem String der Länge  $n$  gibt es höchstens  $n$  Teilworte, die maximale Repeats sind.*

## Aufgabe 2 (Verallgemeinerter Suffixbaum)

(4 Punkte)

Gegeben seien die Strings  $s = \text{GAAG}$  und  $t = \text{AAGG}$ .

1. Zeiche den generalisierten Suffixbaum von  $s$  und  $t$  (mit  $\# < \$ < \mathbf{A} < \mathbf{G}$ ).
2. Gib das längste gemeinsame Teilwort von  $s$  und  $t$  an.
3. Formuliere, wie man allgemein längste gemeinsame Teilworte zweier Strings im generalisierten Suffixbaum finden kann.
4. Ein Student möchte das längste palindromische Teilwort eines Wortes  $x$  finden. Seine Idee: Er betrachtet den generalisierten Suffixbaum von  $x\#\overleftarrow{x}\$$  und nimmt das längste gemeinsame Teilwort von  $x$  und  $\overleftarrow{x}$  als Antwort. Für  $x = \text{BANANAS}$  funktioniert das, aber klappt es immer?

## Aufgabe 3 (Maximal Unique Matches)

(2 Punkte)

Gegeben seien die Strings  $s = \text{BANANENBAUM}$  und  $t = \text{BAUMMANA}$ . Entscheide für jedes der folgenden Wörter, ob es ein MUM von  $s$  und  $t$  ist. Wenn ja, gib die Position der MUMs in  $s$  und  $t$  an (die Indizierung beginne bei 0). Ansonsten begründe, warum es sich nicht um ein MUM handelt.

1. BAUM
2. AN
3.  $\varepsilon$
4. NEN