

# Übungen zur Vorlesung Grundlagen der Sequenzanalyse

Universität Bielefeld, WS 2006/07

Dr. Sven Rahmann · Dipl.-Bioinf. Katharina Jahn

<http://gi.cebitec.uni-bielefeld.de/teaching/2006winter/sequenzanalyse/>

**Blatt 11 vom 11.01.2007**

**Abgabe am 18.01.2007 vor der Vorlesung um 8:30 in H3**

**Aufgabe 1** Gegeben ist  $s = \text{ABACACABA}$ . Gib das Suffixarray  $\text{pos}$ , sein Inverses  $\text{rank}$  und das  $\text{lcp}$ -Array von  $s\$$  an. Beachte  $\$ < A < B < C$ .

**Aufgabe 2** Inwiefern ist das Suffixarray  $\text{pos}$  ein  $q$ -gram Index für alle  $q \geq 1$  gleichzeitig?

**Aufgabe 3** Gib jeweils einen Algorithmus an, der alle längsten wiederholten Substrings in einem String findet,

- (a) mit Hilfe der Tabellen eines Suffixarrays
- (b) mit Hilfe eines Suffixbaums

**Aufgabe 4** In der Vorlesung wurde ein Algorithmus besprochen, mit dem man das  $\text{lcp}$ -Array in linearer Zeit aus  $\text{pos}$  (und seinem Inversen  $\text{rank}$ ) konstruieren kann. Veranschauliche die einzelnen Schritte, in denen dieser Algorithmus auf  $s\$$  aus Aufgabe 1 arbeitet.