

# 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 7 vom 25.05.2012

Abgabe in einer Woche vor Beginn der Vorlesung.

### Aufgabe 1 (Burrows-Wheeler Transformation)

(4 Punkte)

Gegeben sei der String  $t = s\$ = ATGTGATGGTGA\$$ .

1. Was ist die Burrows-Wheeler Transformation? Beschreibe ihre zentrale Idee in Bezug auf die Eigenschaft natürlicher Sprache.
2. Berechne die Burrows-Wheeler Transformierte  $bwt(t)$  für  $t$ .
3. Schreibe  $bwt(t)$  als komprimierten String mit Hilfe von *run-length encoding* auf. Fasse dabei nur Buchstaben zusammen, die min. 3 mal hintereinander vorkommen.

### Aufgabe 2 (Burrows-Wheeler Retransformation)

(5 Punkte)

Gegeben sei die *run-length encoded* Burrows-Wheeler Transformierte  $rle(bwt(t)) = T\$4TG4A$ . Überprüfe, ob das Muster  $p = ATGT$  im Text  $t$  vorkommt. Gehe wie folgt vor:

1. Schreibe  $bwt(t)$  ohne *run-length encoding* auf.
2. Rekonstruiere  $t$  durch Retransformation von  $bwt(t)$ .
3. Beschreibe den String Matching Algorithmus unter Verwendung der  $bwt(t)$  beispielhaft an der Suche von  $p$ .

### Aufgabe 3 (Forward-Backward Technik)

(6 Punkte)

Gegeben seien die beiden Strings  $x = CCATG$  und  $y = TAT$ . Die Indizierung beginnt bei 1.

1. Berechne die Matrizen  $D, D^{rev} = D^{-1}$  und  $C$  unter dem Einheitskostenmodell.
2. Lies alle optimalen Alignments aus  $C$  ab und gib diese an.
3. Welche minimalen zusätzlichen Kosten hat ein Alignment unter der Einschränkung, dass  $x[3]$  und  $y[1]$  aligniert sind.
4. Gib alle Alignments an, die im Vergleich zum optimalen Alignment zusätzliche Kosten von 1 haben.