

Übungen zur Vorlesung Sequenzanalyse

Universität Bielefeld, WS 2015/2016

Prof. Dr. Jens Stoye · M.Sc. Linda Sundermann

<http://wiki.techfak.uni-bielefeld.de/gi/Teaching/2015winter/SequenzAnalyse>

Übungsblatt 9 vom 12.01.2016

Abgabe in einer Woche vor Beginn der Vorlesung.

Aufgabe 1 (Burrows-Wheeler Transformation)

(4 Punkte)

Gegeben sei der String $s = CTGTGTTG\$$.

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

Aufgabe 2 (String Matching mit BWT)

(5 Punkte)

Gegeben sei die *run-length encoded* Burrows-Wheeler Transformierte $rle(bwt(s)) = 2G\$G2ACA2C$.

1. Überprüfe, ob das Muster $p = ACG$ im Text s vorkommt. Gehe wie folgt vor:
 - (a) Dekodiere $rle(bwt(s))$, d.h. schreibe $bwt(s)$ ohne *run-length encoding* auf.
 - (b) Beschreibe den String Matching Algorithmus unter Verwendung der $bwt(s)$ beispielhaft und schrittweise an der Suche von p .
2. Erkläre, wie du erkennen kannst, wie häufig das Muster $p' = CG$ im Text s vorkommt.

Aufgabe 3 (Distanzen und Metriken)

(4 Punkte)

Gegeben seien die zwei Strings $x = ATTT$ und $y = ATTA$. Berechne

- die Hamming-Distanz $d_H(x, y)$,
- die q -gram-Distanz $d_2(x, y)$ für $q = 2$,
- die Maximal-Matches-Distanzen $\delta(x||y)$ und $\delta(y||x)$,
- die Distanz mit der Metrik, die aus der Maximal-Matches-Distanz hervorgeht, $d_{||}(x, y)$ (siehe Theorem 3.25, Seite 26).