

Übungen zur Vorlesung Sequenzanalyse

Universität Bielefeld, WS 2014/2015

Dr. Roland Wittler · Nina Luhmann · Linda Sundermann

<http://wiki.techfak.uni-bielefeld.de/gi/Teaching/2014winter/SequenzAnalyse>

Übungsblatt 8 vom 02.12.2014

Abgabe in einer Woche vor Beginn der Vorlesung.

Aufgabe 1 (Manber-Myers Algorithmus)

(4 Punkte)

1. Führe den Manber-Myers Algorithmus schrittweise für den String $s = \text{ATACAATCTCTAT}$ aus und gib das Suffixarray pos für s an.
2. Wie viele Phasen braucht man für einen String der Länge 13 maximal?

Aufgabe 2 (Suffixarray)

(4 Punkte)

Gegeben ist der String $s = \text{ATACAATCTCTAT}$ aus Aufgabe 1. Beachte $\$ < A < C < G < T$ und Indizierung beginnend mit 0.

1. Implementiere zwei Funktionen, die das $rank$ - und lcp -Array berechnen, wenn das Array pos gegeben ist. Die Funktionen sollen so in ein Programm eingebettet sein, dass der Benutzer nur das pos -Array übergeben muss.
Verwende eine Programmiersprache, die mit deinem Tutor abgesprachen ist und sende ihm deinen Quellcode per Email zu.
2. Wie lauten das $rank$ - und lcp -Array für s ? Gib einen Beispielaufruf für dein Programm mit dem pos -Array von s an.

Aufgabe 3 (Burrows-Wheeler Transformation)

(4 Punkte)

Gegeben sei der String $t = \text{ATGTGATGGTGA}\$$.

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 $rle(bwt(t))$ als komprimierten String mit Hilfe von *run-length encoding* auf. Fasse dabei nur Buchstaben zusammen, die mindestens 3 mal hintereinander vorkommen.

Aufgabe 4 (String Matching mit BWT)

(4 Punkte)

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

1. Dekodiere $rle(bwt(t))$, d.h. schreibe $bwt(t)$ ohne *run-length encoding* auf.
2. Beschreibe den String Matching Algorithmus unter Verwendung der $bwt(t)$ beispielhaft an der Suche von p .