

Übungen zur Vorlesung Sequenzanalyse I

Universität Bielefeld, Wintersemester 2010/2011
Dipl.-Inform Peter Husemann · Dr. Roland Wittler

<http://wiki.techfak.uni-bielefeld.de/gi/Teaching/2010winter/SequenzAnalyse>

Blatt 6 vom 19.11.2010

Abgabe in einer Woche vor Beginn der Vorlesung.

Aufgabe 1 De Bruijn Subgraph

(3 Punkte)

Gegeben sei das Wort $x = \text{ATCCTCTTCA}$ und $q = 3$.

1. Zeichne den De Bruijn Subgraphen $B(x, q)$. (Labele darin die Kanten mit ihrem entsprechenden q -Gram)
2. Nutze den Graphen $B(x, q)$, um das q -Gram Profil von x aufzuschreiben. (Schreibe die q -Gramme dabei in lexikographischer Ordnung auf)
3. Wie viele verschiedene Sequenzen y haben eine q -Gram-Distanz $d_q(x, y) = 0$, obwohl $x \neq y$ ist? Zähle *alle* Sequenzen auf, die das gleiche q -Gram Profil wie x haben.

Aufgabe 2 Partitionen

(3 Punkte)

Gib für die folgenden Sequenzen x und y jeweils

- eine möglichst kurze Partition von x bezüglich y , sowie
- eine möglichst kurze Partition von y bezüglich x

an: (Unterstreiche die b_i der Partitionen)

1. $x = \text{BABBABBABAAAA}$ und $y = \text{BAA}$
2. $x = \text{ACA}$ und $y = \text{BAAB}$
3. $x = \text{ABC}$ und $y = \varepsilon$

Aufgabe 3 Maximal-Matches Distanz

(4 Punkte)

Berechne die *Maximal-Matches Distanz* δ und gib jeweils die entsprechende *links-nach-rechts* (P_{r}) sowie die *rechts-nach-links* Partition (P_{r}) an:

1. $\delta(\text{ANANAS}||\text{ANNASHASE})$
2. $\delta(\text{ANNASHASE}||\text{ANANAS})$