

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 11 vom 14.01.2011

Abgabe in einer Woche vor Beginn der Vorlesung.

Aufgabe 1 Dot Plot

(3 Punkte)

Zeichne einen *Dot Plot* für $x = \text{FEUERREGEN}$ und $y = \text{FEGEFUEER}$. Markiere die Treffer, die nach einer Filterung mit $q = 2$ bzw. $q = 4$ noch sichtbar sind.

Aufgabe 2 Q-Gramm-Lemma

(3 Punkte)

Beweise das Q-Gramm-Lemma. Hinweise: Wie viele Q-Gramme wären in einem Alignment enthalten, wenn die Sequenzen identisch wären? Wie viele Q-Gramme kann ein Mismatch höchstens "zerstören"? Wie viele Q-Gramme bleiben bei e Fehlern also mindestens übrig?

Aufgabe 3 FASTA

(4 Punkte)

1. Erstelle per Hand einen Q-Gramm-Index für $x = \text{AABBABA}$ mit $q = 2$.
2. Gib die Arrays *first* und *pos* an.
3. Schiebe ein Fenster der Länge 2 über $y = \text{ABABBAAB}$ und verwende den Q-Gramm-Index, um $c(x, y; d)$ für alle $d \in \{-|x| \dots |y|\}$ zu berechnen.
4. Wie ist der FASTA-Score $C(x, y)$?

Aufgabe 4 BLAST

(5 Punkte)

Gegeben sei die Query $x = \text{PAPALAPAP}$ und $q = 4$. Der Score für einen Match sei +3 und für einen Mismatch -1. Betrachte nun die Nachbarschaft von x :

1. Gib für $i = 3$ und $i = 5$ jeweils zwei Tupel (z, i) an, die sich in der Nachbarschaft $N_8(x)$ befinden, aber nicht in $N_9(x)$.
2. Bestimme die folgenden Listen $P_k(z)$ (Beachte die verschiedenen Werte für $k!$):
 - (a) $P_4(\text{AAAA}), P_8(\text{AAAA})$
 - (b) $P_8(\text{PAPA}), P_9(\text{PAPA})$
 - (c) $P_8(\text{LALA}), P_9(\text{LALA})$