

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 1 vom 13.04.2012

Abgabe in einer Woche vor Beginn der Vorlesung.

Organisatorisches

Wichtige Hinweise: Bitte unbedingt im eKVV für die Vorlesung (und die Übungen) registrieren. Auf der Homepage der Veranstaltung steht eine elektronische Version des Skripts zum Download bereit. Die gedruckte Version sollte aus dem letzten Semester vorhanden sein. Wer jedoch noch kein aktuelles Skript haben sollte, gibt seinem Tutor in der ersten Übung bescheid.

Abgabe der Übungszettel und Besuch der Tutorien: Die Abgabe der Übungszettel und der Besuch der Tutorien ist im Sommersemester freiwillig. Es müssen keine Aufgaben gelöst werden, um zur Klausur zugelassen zu werden. Trotzdem wird jedem empfohlen, an den Tutorien teilzunehmen. Durch regelmäßige Teilnahme (mindestens zwei Mal Vorrechnen und aktive Mitarbeit) und ein Erreichen von mindestens 50 % der möglichen Punkte, können zwei Creditpoints für die individuelle Ergänzung erlangt werden. Die Aufgaben dürfen dazu wieder in einer Gruppe von maximal zwei Personen bearbeitet werden, müssen aber separat unter eigenem Namen abgeben werden. Die Abgabe erfolgt **vor** Beginn der Vorlesung, entweder beim Veranstalter oder im Briefkasten bei Raum U10-151. Bitte den Tutorientermin bzw. den Tutor, sowie euren Namen deutlich auf der Abgabe vermerken.

Quellenangabe auf den Übungszetteln: Benutzt ihr bei der Bearbeitung der Übungsaufgaben andere Quellen als das Skript zur Vorlesung, so gebt diese bitte immer mit an. Ist nicht ersichtlich, woher ihr eure Informationen genommen habt, können wir euch leider keine Punkte geben.

Teilnahme an der Klausur: Die Klausur kann am Ende jeder mitschreiben, der bis dahin A&D bestanden hat.

Aufgaben

Aufgabe 1 (FASTA)

(4 Punkte)

1. Erstelle per Hand einen Q-Gramm-Index für $x = \text{ATTAAT}$ mit $q = 2$.
2. Gib die Arrays *first* und *pos* an.
3. Schiebe ein Fenster der Länge 2 über $y = \text{ATTTATT}$ 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 2 (BLAST)

(3 Punkte)

Informiere Dich genauer über BLAST.

1. Welche Rolle spielt der **X-Drop Algorithmus** und wie funktioniert er genau?
2. Was sind die Unterschiede zwischen BLAST Version 1.4 und den Versionen ab 2.0?
3. Was bedeutet der Begriff *Sensitivität* im Kontext eines Matches?

Bitte wenden!

Aufgabe 3 (Q-Gramm Statistik)

(2 Punkte)

Seien X und Y zwei Zufallssequenzen nach dem i.i.d.-Modell. Betrachte nun die zwei Q-Gramme $X[i \dots i + q - 1]$ und $Y[j \dots j + q - 1]$. Gib die Formeln für die folgenden Wahrscheinlichkeiten an und erläutere jeweils deine Herleitung.

1. Wie hoch ist die Wahrscheinlichkeit, dass die Q-Gramme **genau ein** Mismatch enthalten?

$$\mathbb{P}(d_H(X[i \dots i + q - 1], Y[j \dots j + q - 1]) = 1) = \dots$$

2. Wie hoch ist die Wahrscheinlichkeit, dass die Q-Gramme **höchstens ein** Mismatch enthalten?

$$\mathbb{P}(d_H(X[i \dots i + q - 1], Y[j \dots j + q - 1]) \leq 1) = \dots$$