

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 1 vom 14.10.2014

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 wird in der ersten Übung verteilt.

Abgabe der Übungszettel: Auch wenn die Bearbeitung der Aufgaben in einer Gruppe (maximal zwei Personen) erfolgt ist, so müssen die Lösungen zu den Aufgaben separat und unter eigenem Namen **vor** Beginn der Vorlesung am Dienstag im Hörsaal abgegeben werden. Alternativ könnt ihr die Zettel auch euren Tutoren geben. Bitte den Tutorientermin bzw. den Tutor, sowie euren Namen und den eures Gruppenpartners 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 mitschreiben, wer bis dahin A&D, Grundlagen der Programmierung oder Ähnliches bestanden, mindestens 50% der Übungspunkte erreicht und mindestens zweimal in den Übungen eine Aufgabe vorgerechnet hat.

Aufgaben

Aufgabe 1 (Minimale und maximale Hammingdistanz) (4 Punkte)

Bei der Translation werden Aminosäuren durch Basentriplets codiert.

Beispiel: Methionin wird nur durch AUG codiert, Arginin hingegen wird durch CGU, CGC, CGA, CGG, AGA und AGG codiert.¹

1. Gegeben seien zwei codierende DNA-Sequenzen der Länge $n = 3m$; ihre Hamming-Distanz betrage $h \in [0, n]$. Wie groß kann die Hamming-Distanz k der zugehörigen Aminosäuresequenzen der Länge m minimal und maximal sein? Gib einfache Sequenzbeispiele für die Extremfälle an.
2. Jetzt seien zwei Aminosäuresequenzen der Länge m mit Hammingdistanz k gegeben. Wie groß kann die Hammingdistanz h von zugehörigen codierenden DNA-Sequenzen der Länge $n = 3m$ minimal und maximal sein? Gib wiederum Beispiele an.

Aufgabe 2 (Implementation der Hamming-Distanz) (4 Punkte)

In der Vorlesung wurde die Hamming-Distanz eingeführt. Implementiere eine Funktion, die als Parameter zwei Strings a und b übergeben bekommt und ihre Hamming-Distanz ausgibt. Wenn die beiden Strings eine unterschiedliche Länge haben, soll eine Fehlermeldung ausgegeben werden. Dein Tutor spricht mit dir ab, welche Programmiersprachen in Ordnung sind.

Aufgabe 3 (Edit-Operationen) (4 Punkte)

Berechne den String $S = E(NXWZLOKP, CS_E CDS_Y CI_R CD)$, gib alle Zwischenschritte an. Wie hoch sind die Kosten der Edit-Sequenz? Gib eine weitere Edit-Sequenz mit maximalen Kosten an, die den gleichen String S ergibt.

¹Mehr dazu unter: http://de.wikipedia.org/wiki/Genetischer_Code

Aufgabe 4 (Berechnung der Edit-Distanz)

(3 Punkte)

Gegeben seien die Strings $x = \text{FLAMINGO}$ und $y = \text{BINGO}$. Was ist die Edit-Distanz der beiden Strings? Benutze zur Berechnung eine Edit-Matrix.