

4. Übung zur Vorlesung “Einführung in die Bioinformatik I, 2. Teil”

Sommersemester 2020

Prof. Peter Dittrich, Marcus Ludwig

Ausgabe: 26. Mai 2020,
Abgabe: 02. Juni 2020 per Email

Aufgabe 1 (3 Punkte): Erklären Sie den Unterschied zwischen der Kante (u, v) in einem gerichteten Graph und der Kante $\{u, v\}$ in einem ungerichteten Graph? Was ändert sich, wenn man u und v vertauscht? Ist $u = v$ erlaubt?

Aufgabe 2 (6 Punkte): Gegeben sind die vier Sequenzen $s_1 = GAA$, $s_2 = GACA$, $s_3 = ACAC$ sowie $s_4 = CACG$ und der Leitbaum $((1, 2), (3, 4))$. Bestimmen Sie das multiple globale Alignment der vier Sequenzen mit Hilfe des Feng-Doolittle-Verfahrens für die Ähnlichkeitsfunktion $S(A, A) = 2$, $S(a, a) = 1$ für alle $a \neq A$, $S(a, b) = -1$ für $a \neq b$ und $S(a, -) = S(-, a) = -1$ für alle a .

Aufgabe 3 (6 Punkte): Überlegen Sie sich ein Beispiel für ein progressives globales Alignment mit drei Sequenzen, bei dem sich mit dem Feng-Doolittle-Verfahren je nach Leitbaum unterschiedliche multiple Alignments ergeben.

Aufgabe 4 (5 Punkte): Analysieren Sie Laufzeit und Speicherbedarf des Feng-Doolittle-Verfahrens bei Eingabe von k Sequenzen der Länge n und gegebenem Leitbaum. Annahme: Die Sequenzen dürfen sich während des Progressiven Alignments nur um einen konstanten Faktor verlängern (Das bedeutet, selbst mit eingefügten Gaps ist die Länge einer Sequenz immer $O(n)$).