

Klausur zum Modul

Algorithmen in der Bioinformatik

Nachklausur Sommersemester 2018
19.09.2018

Name:

Matrikelnummer:

Studiengang:

Geben Sie den Lösungsweg immer mit an!

Nur mit blauem oder schwarzem Kugelschreiber schreiben.

Schreiben Sie auf jeden Zettel Ihre Matrikelnummer.

Geben Sie für jede (Teil-)Aufgabe nur eine einzige Lösung ab. Bei mehreren, alternativen Lösungen zu einer Aufgabe wird die Schlechteste bewertet.

Teilnahme an der Klausur erfolgt unter Vorbehalt einer vorhandenen Zulassung.

Aufgabe Nr.:	Punktzahl:	Davon erreicht:
1	15	
2	15	
3	10	
4	10	
Σ	50	

Es sind keinerlei Hilfsmittel erlaubt. Bitte schreiben Sie deutlich mit einem schwarzen oder blauen Stift.

1. Alignments

- a) Geben Sie die Dynamische Programmierungsmatrix mit Backtrackingpointern an, um optimale globale Alignments zwischen den Sequenzen **AAAC** und **AGC** zu bestimmen. Ein Match wird mit $+1$ bewertet, ein Mismatch wird mit -1 bestraft und ein Gap wird mit -2 bestraft.

9

Geben Sie die optimalen Alignments und deren Scores an.

- b) Wie muss man den Algorithmus für lokales Sequenzalignment ab ändern, wenn der längste gemeinsame Substring der Inputsequenzen berechnet werden soll?

6**2. Sortieren durch Umkehrungen**

Gegeben ist die Sequenz **6 1 2 3 4 5**.

- a) Wenden Sie den gierigen Algorithmus **SimpleReversalSort** auf diese Sequenz an. Zeigen Sie alle Umkehrungsschritte. Wieviele Schritte werden ausgeführt?
- b) Was ist die maximale Anzahl von Schritten, die **SimpleReversalSort** benötigt, um eine Permutation der Länge n zu sortieren.
- c) Analysieren Sie die Worst-Case-Laufzeit von **SimpleReversalSort**.

5**5****5****3. Assembly**

Gegeben ist der String **HOLTERDIPOLTER**.

- a) Zeichnen Sie den De-Bruijn-Graphen G auf Basis aller 3-mere des Strings.
- b) Finden Sie einen Eulerweg in G . Wieviele verschiedene Eulerwege gibt es in dem Graphen? Zu welchen Strings korrespondieren diese Wege?

5**5****4. Clustering**

- a) Welches Maß soll beim k -Means-Clustering optimiert werden?
- b) Geben Sie eine Beispielinstanz und Startzentren für k -Means-Clustering mit dem Lloyd-Algorithmus an, für die der Algorithmus keine Optimierung berechnet.

5**5**