
Algorithmische Bioinformatik II

Abgabetermin: Freitag, den 13. Dezember, 12⁰⁰

Tutoraufgabe 1 (Vorbereitung bis zum 11.12.19)

Zeige, dass es für jedes $n \in \mathbb{N}$ eine Menge von Sequenzen $S = \{s_1, \dots, s_k\}$ mit $|s_i| \geq n$ gibt, so dass es eine Sequenz $s \in S$ gibt, die als Zentrum bei der Center-Star-Methode eine Approximationsgüte von $\Omega(k)$ liefert. Hierbei gilt $w(x, y) = 1$ und $w(x, x) = 0$ für alle $x \neq y \in \bar{\Sigma}$.

Hinweis: Die Menge ist hier als Multimenge zu verstehen, d.h. Sequenzen dürfen mehrfach in S vorkommen.

Aufgabe (Notenbonus) 2

Betrachte die Erweiterung eines zu einem Baum konsistenten Alignments wie im Induktionsschritt des Beweises von Satz 6.35 im Skript. Gib eine konkrete Datenstruktur für mehrfache Alignments und die zugehörige Implementierung der Erweiterung von Alignments an, so dass diese die gewünschte Laufzeit von $O(kn)$ eingehalten wird. Analysiere den Zeitbedarf Deiner Implementierung möglichst genau.

Aufgabe (Notenbonus) 3

Betrachte folgende Sequenzen $s_1 = CGAA$, $s_2 = CAGTGA$, $s_3 = CAATG$ und $s_4 = CGGATT$. Der optimale Abstand für die paarweise Sequenzen-Alignments ist rechts angegeben. Hierbei gilt $w(a, b) = 1$ und $w(a, a) = 0$ für alle $a \neq b \in \bar{\Sigma}$. Konstruiere für diese Sequenzen ein mehrfaches Sequenzen-Alignment mit Hilfe der Center-Star-Methode aus der Vorlesung.

d	s_1	s_2	s_3	s_4
s_1	0	3	3	3
s_2	3	0	2	4
s_3	3	2	0	3
s_4	3	4	3	0