
Algorithmische Bioinformatik II

Abgabetermin: Freitag, den 12. Dezember, 9⁰⁰ Uhr in Moodle

Hausaufgabe 1

Betrachte die Erweiterung eines zu einem Baum konsistenten Alignments wie im Induktionsschritt des Beweises von Lemma 6.35 im Skript. Gib eine konkrete Datenstruktur für mehrfache Sequenzen-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.

Hausaufgabe 2

Betrachte folgende Sequenzen $s_1 = CAGC$, $s_2 = CGTGTC$, $s_3 = CGCGT$ und $s_4 = CGACGT$. 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	3
s_3	3	2	0	1
s_4	3	3	1	0

Tutoraufgabe 3 (Vorbereitung bis zum 10.12.25)

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.