
Algorithmische Bioinformatik II

Abgabetermin: Freitag, den 16. Dezember, 9⁰⁰ Uhr in Moodle

Tutoraufgabe 1 (Vorbereitung bis zum 14.12.22)

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.

Hausaufgabe 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 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 3

Betrachte die Sequenzen $s_1 = \text{AGACA}$, $s_2 = \text{GACAC}$ und $s_3 = \text{ACCA}$. Berechne die C -optimalen Schnittpositionen mit Respekt zu $c_1 = 1$ sowie zu $c_1 = 2$ und die daraus resultierenden mehrfachen Alignments gemäß des Divide-and-Conquer-Alignment-Algorithmus, wobei nach der ersten Rekursion bereits jeweils ein optimales Alignment für die jeweiligen Präfixe bzw. Suffixe berechnet wird. Die zugrunde liegende Kostenfunktion für das SP-Distanzmaß sei $w(a, a) = 0$ und $w(a, b) = 1$ für alle $a \neq b \in \bar{\Sigma}$.