Ludwig-Maximilians-Universität München
Institut für Informatik
Prof. Dr. Volker Heun

Wintersemester 2010/11
Übungsblatt 6
2. Dezember 2010

Algorithmische Bioinformatik II

Abgabetermin: Donnerstag, den 9. Dezember, vor der Vorlesung

Aufgabe 1

Beweise oder widerlege: Ein optimales mehrfaches Sequenzen-Alignment bezüglich des Sum-of-Pair-Maßes induziert mindestens ein optimales paarweises Sequenzen-Alignment.

Aufgabe 2

Betrachte die Sequenzen $s_1 = \text{AAGCC}$, $s_2 = \text{CCTGG}$ und $s_3 = \text{GGCAA}$. Berechne die C-optimalen Schnittpositionen mit Respekt zu $c_1 = 1$ sowie $c_1 = 2$ 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 $a \neq b \in \overline{\Sigma}$.

Aufgabe 3

Betrachte die Erweiterung eines zu einem Baum konsistenten Alignments wie im Induktionschritt des Beweises von Satz 8.35 im Skript. Gib eine konkrete Datenstruktur für mehrfache Alignments und die zugehörige Implementierung der Erweiterung von Alignments an. Analysiere den Zeitbedarf hierfür möglichst genau.

Aufgabe 4

Betrachte eine Spalte bei der Erweiterung eines zu einem Baum konsistenten Alignments. Angenommen im optimalen paarweisen Alignment wurde ein Leerzeichen in die Sequenz eingeführt, die bereits im konstruierten mehrfachen Alignment enthalten ist und die an dieser Position dort im mehrfachen Alignment bereits ein Leerzeichen enthält. Es gibt nun zwei Möglichkeiten, diese Insertion im paarweisen Alignment in das mehrfache Alignment zu integrieren. Entweder man verwendet das Leerzeichen wieder (und tut eigentlich gar nichts) oder man fügt ein weiteres Leerzeichen in das mehrfache Alignment ein und fügt somit eine Spalte $(-, \ldots, -, a)$ für die Insertion $(-, a)$ des paarweisen Alignments ein.

Zeige, dass es unter vernünftigen Annahmen über die Kostenfunktion einer Distanzfunktion für ein paarweises Alignment (welche!) zu einer geringeren SP-Distanz führt, wenn die erste Konstruktion ausgeführt wird.