Technische Universität München
Institut für Informatik
Lehrstuhl für Effiziente Algorithmen
Volker Heun

SS 2000
Übungsblatt 3
23. Mai 2000

Algorithmen der Bioinformatik

Abgabetermin: 30. Mai

Aufgabe 1

a) In der Vorlesung wurde eine Methode zur Konstruktion eines *verallgemeinerten Suffixbaums* zu einer Menge $M = \{w_1, w_2, \ldots, w_n\}$ von Strings vorgeführt. Dabei wurde ein künstliches Wort $w' = w_1\$_1 w_2\$_2 \cdots w_n\$_n$ betrachtet. Dieser Baum ist jedoch größer als notwendig. Geben Sie eine Methode zur Erzeugung eines kleineren Suffixbaums an, der alle Suffixe der Wörter von M enthält.

b) Konstruieren Sie nach Ihrer Methode den Suffixbaum für die Sprache

$$\{0101100011, 1101, 01010\}.$$

Aufgabe 2

Sei, wie in der Vorlesung definiert, $w : (\Sigma \cup \{-\})^2 \longrightarrow \mathbb{R}_0^+$ die Kostenfunktion für Edit-Operationen eines Sequence Alignment, und seien die Gesamtkosten für eine Folge $S = s_1 s_2 \ldots s_n$ von Edit-Operationen definiert als $d_w(S) = \sum_{i=1}^{n} w(s_i)$.
Zeigen Sie: Ist die Funktion w eine Metrik, so ist auch d_w eine Metrik.

Aufgabe 3

Im WWW finden Sie unter www14.in.tum.de/lehre/2000SS/cb/notes/ oder unter www.cs.arizona.edu/people/gene/ den Technical Report "An Overview of Sequence Comparison Algorithms in Molecular Biology" von E. Myers, 1991. Lesen Sie den Bericht, so daß er in der Übung diskutiert werden kann.