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

SS 2000
Übungsblatt 6
27. Juni 2000

Algorithmen der Bioinformatik

Abgabetermin: 4. Juli

Aufgabe 1

In der Stern-Methode zur approximativen Konstruktion eines mehrfachen Sequenzen Alignments wurden beim Hinzufügen der neuen Zeichenreihen zu dem bereits existierenden MSA keine Bedingung an die Reihenfolge vorgegeben. Zeigen Sie, dass unabhängig von der Reihenfolge immer dasselbe mehrfache Sequenzen Alignment konstruiert wird.

Aufgabe 2

Sei $S = \{s_1, \ldots, s_k\}$ eine Menge von Strings, so dass kein String Teilstring eines anderen Strings ist. Zeigen Sie, dass für einen kürzesten Superstring SSP(S) gilt:

$$\max\{|s_i| \,:\, i \in [1:k]\} + k - 1 \;\leq\; \mathrm{SSP}(S) \;\leq\; \sum_{i=1}^{k} |s_k|.$$

Geben Sie Beispiele für beliebiges k an, die zeigen, dass die Schranken scharf sind.

Aufgabe 3

Sei $S = \{s_1, \ldots, s_k\}$ eine Menge von Strings. Zeigen Sie, dass Sie in Zeit $O(\sum_{i=1}^{k} |s_i|)$ die folgende Menge bestimmen können:

$$\{s \in S \,:\, \exists i \in [1:k] : s \text{ ist ein Teilwort von } s_i\}.$$

Dies ist die Menge aller Strings, die in einem anderen String enthalten sind.

Hinweis: Verwenden Sie Suffix-Bäume.