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

SS 2000
Übungsblatt 2
30. Mai 2000

Algorithmen der Bioinformatik

Abgabetermin: 23. Mai

Aufgabe 1

Geben Sie den Suffixbaum für das Wort $T = acctagctagct$ an. Ist dieser eindeutig?

Aufgabe 2

a) Geben Sie einen Algorithmus für das Problem *exact string matching* auf der Basis von Suffixbäumen an, der zu einem Text T und einem Suchwort P die Position *jedes* Vorkommens von P in T ausgibt. Für k Vorkommen von P in T soll die Komplexität des Algorithmus $\mathrm{O}(|T| + |P| + k)$ betragen.

b) Wie kann man die Komplexität der Suche nach P auf $\mathrm{O}(|P|)$ senken, wenn nur höchstens ein Vorkommen von P in T auszugeben ist?

c) Welchen Platz belegt eine optimal implementierte Datenstruktur für einen Suffixbaum eines Textes der Länge n?

Aufgabe 3

Gegeben sei eine Menge $M = \{s_1, \ldots, s_n\}$ von Strings mit $n = \sum_{i=1}^{n} |s_i|$. Geben Sie einen Algorithmus an, mit dem man in Zeit $\mathrm{O}(n)$ alle Strings $s_i \in M$ finden kann, die Teilstrings eines anderen Strings $s_j \in M$ sind.