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

Wintersemester 2010/11
Übungsblatt 9
24. Juni 2010

Algorithmische Bioinformatik I

Abgabetermin: Donnerstag, den 1. Juli, vor der Vorlesung

Aufgabe 1

Konstruiere einen Suffix-Baum für $t = t_1 \cdots t_{10} = aaabaababb$ mit Ukkonens Algorithmus aus der Vorlesung. Gib dabei alle Zwischenschritte an, markiere insbesondere die Position des aktiven Knotens und Endknotens im jeweiligen Suffix-Baum. Zeichne dabei nur die verwendeten und neu eingetragenen Suffix-Links mit jeweils einer anderen Farbe ein. Der Übersichtlichkeit wegen können als Kantenlabel die Zeichenreihen selbst anstelle der üblicherweise verwendeten Referenzen auf t verwendet werden.

Aufgabe 2

Entwirf einen effizienten Algorithmus, der in einem Text t das längste Teilwort w findet, das zweimal in t als Teilwort auftritt.

Aufgabe 3

Ein Suffix-Array S für ein Wort $t \in \Sigma^n$ ist Feld der Größe n mit folgenden Eigenschaften:

i) $\forall i < j \in [1:n]: \; t_{S[i]} \cdots t_n < t_{S[j]} \cdots t_n$,

ii) $\{S[i] \,:\, i \in [1:n]\} = [1:n]$.

Gib einen Algorithmus mit linear vielen Zeichenvergleichen an, der für t ein Suffix-Array S konstruiert.

Hinweis: Es gilt $x < y$ für $x, y \in \Sigma^*$, wenn es ein $i \in \mathbb{N}$ mit $(x_1, \ldots, x_{i-1}) = (y_1, \ldots, y_{i-1})$ und $x_i < y_i$ (oder $|x| = i - 1 < |y|$) gibt.

Achtung: Ein Vergleich von zwei Suffixen kann schon $O(n)$ Zeichenvergleiche benötigen.

Aufgabe 4

Eine Zeichenkette $w' \in \Sigma^*$ ist eine *zyklische Rotation* einer Zeichenkette $w \in \Sigma^*$, wenn es zwei Zeichenketten $u, v \in \Sigma^*$ gibt, so dass $w = uv$ und $w' = vu$.

Beispiel: BAUMAST ist eine zyklische Rotation von MASTBAU.

Entwirf einen Algorithmus, der für $s \in \Sigma^m$ und $t \in \Sigma^n$ in Zeit $O(n + m)$ feststellt, ob t eine zyklische Rotation von s enthält.