

5. Übung „Algorithmen der Bioinformatik I“

Abgabe am 4. bzw. 5. Juni 2003 in der Übung



Institut für Informatik
Martin-Luther-Universität Halle-Wittenberg

Aufgabe 1. Skizzieren Sie die Konstruktion der impliziten Suffixbäume für den String AGGTAGG über den **linearen** Ukkonen-Algorithmus. Vergessen Sie nicht, die Verwendung der Suffixlinks mit einzubeziehen.

Aufgabe 2. Gegebenen sei ein String S . Wir betrachten den Ukkonen-Algorithmus in Phase $i + 1$ (wir bauen also den impliziten Suffixbaum \mathcal{T}_{i+1} für $S[1..i + 1]$). Die Kantenmarkierungen seien durch Indexpaare (p, q) repräsentiert. Wenn nun in einer beliebigen Erweiterung j der Phase $i + 1$ die Regel I angewendet wird, dann wird das Indexpaar (p, q) der gefundenen Blattkante zu $(p, q+1)$ modifiziert. Zeigen Sie induktiv, dass $q = i$ und die Aktualisierung damit korrekt ist. (Hinweis: Das Problem ist, dass der Teilstring $S[p..q]$ auch an anderer Stelle im String S auftauchen könnte. Die Induktion sollte über die Phasen i gehen.)

Aufgabe 3.

- (i) Zeigen Sie, dass wenn in Phase $i + 1$ des Ukkonen-Algorithmus in einer Erweiterung j die Regel III zutrifft, dann trifft Regel III auch für den Rest der Phase $i + 1$ zu. Was bedeutet das für die Implementierung?
- (ii) Beweisen Sie, dass in jeder Phase $i + 1$ des Ukkonen-Algorithmus **alle** Anwendungen von Regel I am Anfang durchgeführt werden. Zeigen Sie dazu, dass jeder Suffix $S[j..i]$ mit $1 \leq j \leq$ Anzahl der Blätter von \mathcal{T}_i in einem Blatt endet, die Regel I wird also so oft angewendet, wie Blätter im Baum \mathcal{T}_i vorhanden sind.

Aufgabe 4. Der Ukkonen-Algorithmus berechnet zu einem gegebenem String S den zugehörigen impliziten Suffixbaum. Wie kann dieser in einen expliziten Suffixbaum umgewandelt werden?