

6. Übung „Algorithmen der Bioinformatik I“

1. Skizzieren Sie die Konstruktion der impliziten Suffixbäume für den String AGG-TAGG mit Hilfe des Ukkonen-Algorithmus. (2 Punkte)
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 $(p, i + 1)$ modifiziert. Zeigen Sie induktiv, dass $q = i$ gilt 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.) (4 Punkte)
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? (3 Punkte)
(ii) Beweisen Sie, dass in jeder Phase $i + 1$ des Ukkonen-Algorithmus **alle** Anwendungen von Regel I am Anfang durchgeführt werden und dies so oft der Fall ist, wie es Blätter im Baum zu Beginn der Phase $i + 1$ gab. (3 Punkte)

Hinweis zu Teil (ii):

Zeigen Sie dazu zuerst, dass für jedes am Anfang von Phase $i + 1$ vorhandene Blatt die Regel I in Phase $i + 1$ genau einmal angewendet wird und dass es keine weiteren Anwendungen geben kann. Zeigen Sie weiterhin, dass über alle Phasen hinweg die Blattnummern in aufsteigender Reihenfolge und ohne Lücken vergeben werden. Letzteres komplettiert den Beweis der Hauptaussage.