

Algorithmen auf Sequenzen

Übungsblatt 4

Abgabetermin: Dienstag, 22.11.2016, 10 Uhr

(Via email an sophie.friedl@bio.ifi.lmu.de oder persönlich)

1. Aufgabe (Ukkonen) Konstruiere Sie für $t = aaabaabab\$$ einen Suffix-Baum mit Hilfe des Algorithmus von Ukkonen und geben Sie dabei alle Zwischenschritte an.

Hinweis: Für jeden Präfix ist jeweils ein eigener Suffix-Baum zu zeichnen, in dem die neu eingefügten Knoten und Blätter (sowie der verwendete aktive Suffix) zu erkennen sind. Zeichnen Sie dabei die verwendeten Suffix-Links als auch die neu erstellten Suffix-Links jeweils mit einer eigenen Farbe ein.

2. Aufgabe (Ukkonen) Sei $t \in \Sigma^*$ und $k \in \mathbb{N}$ gegeben. Entwerfen Sie einen Algorithmus mit linearer Laufzeit, der alle kürzesten Teilwörter von t findet, die genau k -mal in t auftreten.

Hinweis: Korrektheitsbeweis und Laufzeitanalyse nicht vergessen!

3. Aufgabe (Eindeutige Teilwörter): Gegeben sei eine Zeichenreihe $t \in \Sigma^*$. Ein Wort $w \in \Sigma^*$ heißt *minimal rechts bzw. links-eindeutiges Teilwort* von t , wenn w genau einmal in t auftritt und wenn jedes Präfix bzw. Suffix von w mindestens zwei Mal in t auftritt.

- (a) Entwerfen Sie einen effizienten Algorithmus zum Auffinden aller minimal rechts-eindeutigen Teilwörter der Länge mindestens ℓ , beweisen Sie seine Korrektheit und analysieren Sie seine Laufzeit.
- (b) Entwerfen Sie einen effizienten Algorithmus zum Auffinden aller minimal links-eindeutigen Teilwörter der Länge mindestens ℓ , beweisen Sie seine Korrektheit und analysieren Sie seine Laufzeit.