

Algorithmen auf Sequenzen

Übungsblatt 5

Abgabetermin: Dienstag, 29.11.2016, 10 Uhr
(Via email an sophie.friedl@bio.ifi.lmu.de oder persönlich)

1. Aufgabe (Stoye und Gusfield)

Wenden Sie den Algorithmus von Stoye und Gusfield auf das folgende Wort

$$t = t_1 \cdots t_{13} = \text{abbabbabba}$$

an. Geben Sie dazu für jeden Knoten v seine Blattlisten (getrennt nach $LL(v')$ und $LL'(v)$), sein DFS-Intervall ($\text{DFS_Int}(v)$) sowie das ausgeschlossene DFS-Intervall ($\text{DFS_Int}(v')$) an. Geben Sie weiter für jeden Knoten die ausgeführten Tests (basierend auf den DFS-Intervallen) und deren Ergebnis an (und ggf. das ausgegebene rechtsverzweigende Tandem-Repeat).

2. Aufgabe (Main und Lorentz)

Wenden Sie den Conquer-Step aus dem Algorithmus von Main und Lorentz auf das folgende Wort

$$t = t_1 \cdots t_{20} = \text{abaabbabbabbabbabaa}$$

für $h = \lfloor \frac{n}{2} \rfloor = 10$ und $q = h + \ell$ an (also nur Schritt 3). Geben Sie dazu für jedes $\ell \in [3 : 6]$ die ausgeführten LCE-Anfragen und die ausgegebenen Tandem-Repeat-Paare an (geben Sie zusätzlich an, welche davon rechtsverzweigend sind).

3. Aufgabe (Minimal eindeutiges Teilwort)

Gegeben sei eine Zeichenreihe $t \in \Sigma^*$. Ein Wort $w \in \Sigma^*$ heißt *minimal eindeutiges Teilwort* von t , wenn w genau einmal in t auftritt und wenn jedes Teilwort von w mindestens zweimal in t auftritt.

Entwerfen Sie einen möglichst effizienten Algorithmus zum Auffinden aller minimal eindeutigen Teilwörter der Länge mindestens ℓ , beweisen Sie dessen Korrektheit und analysieren Sie die Laufzeit.