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

Wintersemester 2023/24
Übungsblatt 6
29. November 2023

Algorithmen auf Sequenzen

Abgabetermin: Samstag, den 9. Dezember, 10^{00} in Moodle

Aufgabe 1

Sei $t = t_1 \cdots t_{10} = aabbaaabbb$ und bestimme die maximalen Paare in t der Länge mindestens 2 mithilfe des in der Vorlesung angegebenen Algorithmus basierend auf Suffix-Bäumen.

Gib dazu den zugehörigen annotierten Suffix-Baum $T = T(t\$)$ an, d.h. gib in T die Blattlisten für jeden Knoten an und gib für jeden inneren Knoten $v \in V(T)$ in DFS-Reihenfolge an, welche maximalen Paare an diesem Knoten v ermittelt werden.

Aufgabe 2

Wende den beschleunigten Algorithmus von Stoye und Gusfield (Abb. 3.20 im Skript) auf das folgende Wort

$$t = t_1 \cdots t_{13} = abbabbabbabba$$

an. Gib dazu für jeden Knoten v seine Blattlisten (getrennt nach $LL(v')$ und $LL'(v)$), sein DFS-Intervall (DFS_Int(v)) sowie das DFS-Intervall (DFS_Int(v')) für die ausgewählte längste Blattliste an. Gib weiter für jeden Knoten die ausgeführten Tests (basierend auf den DFS-Intervallen) und deren Ergebnis an (und ggf. das ausgegebene rechtsverzweigende Tandem-Repeat).

Tutoraufgabe 3 (Vorbereitung bis zum 6.12.23)

Beweise das Lemma 3.24 aus der Vorlesung vollständig:

Sei $t \in \Sigma^n$ und sei $i < j \in [1 : n]$ sowie $\ell := j - i > 0$. Dann sind folgende Aussagen äquivalent:

1. *Das Paar (i, ℓ) ist ein rechtsverzweigendes Tandem-Repeat-Paar.*
2. *Es existiert ein Knoten $\overline{v} \in V(T(t\$))$ mit $|v| = \ell$ und $i, j \in LL(\overline{v})$. Weiterhin gilt für alle Knoten $\overline{w} \in V(T(t\$))$ mit $|w| > \ell$, dass nicht sowohl $i \in LL(\overline{w})$ als auch $j \in LL(\overline{w})$ gilt.*