

Text Indexing and Information Retrieval

Übungsblatt 4

Besprechung: 5.12. 2013

Aufgabe 1 (Theorie)

Entwerfen Sie einen Linearzeitalgorithmus, der für einen Text T das *kürzeste* Teilwort findet, das nur einmal in T vorkommt. Hinweis: Suffixbäume!

Aufgabe 2 (Theorie)

Sei T ein Text und T^R der Text in umgekehrter Reihenfolge. Zeigen Sie, dass das LCP-Array für T^R eine Permutation des LCP-Arrays für T ist. Hinweis: Zählen Sie für jedes $\ell \geq 0$ separat die Anzahl der LCP-Array Einträge mit Wert ℓ in beiden LCP-Arrays.

Aufgabe 3 (Theorie)

Entwerfen Sie einen Text-Index linearer Größe, der für ein Muster $P_{1..m}$ ein Array $V[1, m]$ ausgibt, so dass $V[i]$ das längste Präfix von $P_{i..m}$ angibt, das in T vorkommt. Die erwartete Laufzeit soll $O(m)$ sein. Hinweis: Suffixbäume mit entsprechender Zusatzinformation!