

Text Indexing and Information Retrieval

Übungsblatt 6

Besprechung: 28.11.2016

Aufgabe 1 (Theorie und Praxis)

Überlegen Sie sich, wie das LCP-Array als Maß für die Komprimierbarkeit von Texten benutzt werden kann.

Implementieren Sie Ihr Kompressionsmaß und fertigen Sie verschiedene "LCP-Statistiken" für die Texte auf <http://pizzachili.dcc.uchile.cl/texts.html> an.

Aufgabe 2 (Theorie)

Zeigen Sie die komplette Datenstruktur für Level-Ancessor-Anfragen in $O(1)$ Zeit auf dem Suffixbaum für den Text `alabararlabarbarla$`. Für dieses Beispiel setzen Sie den Wert `s` fix auf 4 (statt auf $\frac{\log n}{4}$).

Aufgabe 3 (Theorie)

Zeigen Sie, wie die Vorberechnung der Jump-Pointer in der Level-Ancessor-Datenstruktur in $O(n)$ Zeit geht.

Aufgabe 4 (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.