Algorithmische Grundlagen und Vermittlung der Informatik
Lehrstuhl 11 (Algorithm Engineering)
Prof. Dr. Johannes Fischer

technische universität dortmund

Text Indexing and Information Retrieval

Übungsblatt 6

Besprechung: 17.11.2014

Aufgabe 1 (Praxis)

Implementieren Sie ein LCP-Array, das zunächst nur ein Byte pro Eintrag reserviert. LCP-Werte höher als 254 werden in diesem Array mit '255' markiert und in einer „geeigneten" zweiten Datenstruktur gespeichert. Hierbei sind Ihrer Phantasie keine Grenzen gesetzt.

Implementieren Sie auch den in der Vorlesung vorgestellten LCP-Array Konstruktionsalgorithmus mit dieser neue Darstellung.

Testen Sie Ihre Datenstruktur und Ihren Algorithmus für Texte der Größenordnung 100MB und mehr, insbesondere auf Platz- und Zeitbedarf. Solche Texte können Sie etwa auf der folgenden Seite finden:

`http://pizzachili.dcc.uchile.cl/texts.html`

Aufgabe 2 (Theorie)

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

Aufgabe 3 (Theorie)

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

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.