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

technische universität dortmund

Text Indexing and Information Retrieval

Übungsblatt 3

Besprechung: 21.11. 2013

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.

Passen Sie auch einen in der Vorlesung vorgestellten LCP-Array Konstruktionsalgorithmus auf diese neue Darstellung an.

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)

Zeigen oder widerlegen Sie: wenn im LCP-Array der Wert ℓ (an einer beliebigen Stelle) auftritt, dann tritt auch der Wert $\ell - 1$ (an einer beliebigen anderen Stelle) auf.

Aufgabe 3 (Theorie)

Zeigen Sie:

a) Zu jedem internen Knoten v im Suffixbaum gibt es ein Intervall $[v_\ell, v_r]$, so dass die Werte in $A[v_\ell], \ldots, A[v_r]$ genau die Blattlabels der sich unterhalb von v befindenden Blätter sind.

b) Für das unter (a) genannte Intervall $[v_\ell, v_r]$ gilt: es gibt einen Index $i \in [v_\ell, v_r]$, so dass der LCP-Wert an der Stelle i genau der String-Tiefe von v entspricht.