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

technische universität dortmund

Text Indexing and Information Retrieval

Übungsblatt 10

Besprechung: 5.1.2015

Aufgabe 1 (Theorie)

Finden Sie die LZ77-Zerlegung für den Text $T = \texttt{abaababaabaababaababa}$. Geben Sie auch alle benötigten Datenstrukturen an und zeigen Sie, wie der in der VL gezeigte Algorithmus arbeitet.

Aufgabe 2 (Theorie)

In der VL wurden für die Berechnung der LZ-Zerlegung u.a. alle Datenstrukturen für konstante lcp-Anfragen benötigt (also insbesondere das LCP-Array und die Datenstrukturen für RMQ darauf). Stattdessen könnte man auch auf konstante lcp-Anfragen verzichten, und jeden neuen potentiellen (an Textstelle i beginnenden) Faktor mit den an Positionen $A[PSV_A(i)]$ und $A[NSV_A(i)]$ beginnenden Suffixen *naiv* vergleichen.

Beschreiben Sie dieses Verfahren im Detail und analysieren Sie seine Laufzeit. Zeigen Sie auch, wie das Verfahren auf dem Beispielstring aus Aufgabe 1 arbeitet.

Aufgabe 3 (Theorie)

In der VL wurde auch kurz auf die LZ78-Zerlegung eingegangen. Beschreiben Sie dieses Verfahren im Detail und analysieren Sie seine Laufzeit (hierzu sollten Sie effiziente Trie-Navigation aus der 1. VL nutzen). Für die Analyse der Laufzeit können Sie zudem die Tatsache benutzen, dass für die Anzahl z von Faktoren in der LZ78-Zerlegung die obere Schranke $z \leq n/\log_\sigma n$ gilt (Textlänge n, Alphabetgröße σ).