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

technische universität dortmund

Text Indexing and Information Retrieval

Übungsblatt 2

Besprechung: 31.10.2016

Aufgabe 1 (Praxis)

1. Implementieren Sie einen Algorithmus, der das Suffix Array naiv berechnet, d.h. nutzen Sie einen beliebigen Standard-Sortieralgorithmus (z.B. aus einer Java- oder C++-Bibliothek).
2. Implementieren Sie den in der Vorlesung vorgestellten Algorithmus (Manber, Udi, and Gene Myers. "Suffix arrays: a new method for on-line string searches." *siam Journal on Computing* 22.5 (1993): 935-948.). Sie können auch die vereinfachte Version aus der Vorlesung implementieren. Diese basiert auf der Beschreibung des Manber & Myers Algorithmus in: Puglisi, Simon J., William F. Smyth, and Andrew H. Turpin. "A taxonomy of suffix array construction algorithms." *acm Computing Surveys (CSUR)* 39.2 (2007): 4.
3. Vergleichen Sie die Laufzeiten der beiden Implementierungen. Hierzu können Sie die Texte von `http://pizzachili.dcc.uchile.cl/texts.html` nutzen.

Aufgabe 2 (Praxis)

Führen Sie den Algorithmus aus der Vorlesung für die Texte

- $T_1 =$ `mississippi$` und
- $T_2 =$ `abcdabcd$`

durch. Geben Sie hierzu den Inhalt aller Arrays zu jeden Zeitpunkt an.

Aufgabe 3 (Theorie)

Beim „Prefix Doubling“ werden Präfixe der Länge 2^k betrachtet. Funktioniert diese Technik auch mit Präfixen der Länge 4^k oder sogar α^k für $a \in \mathbb{N}$ mit $\alpha > 1$? Was ist dabei zu beachten?