



Blatt 6

Aufgabe 6.1

- (a) Konstruieren Sie den impliziten Suffixbaum für den String $abbcabbd$ (mit Suffixlinks). Tragen Sie die bei der Konstruktion des Suffixbaumes angewandten Regeln (I, IIa, IIb oder III) in eine Tabelle ein, wobei die Zeilen den Phasen $i + 1 = 1, 2, \dots, n$ und die Spalten den Erweiterungen $j = 1, \dots, i + 1$ entsprechen. Tragen Sie dabei auch die implizit angewandten Regeln ein! (3 Punkte)
- (b) Wir haben bereits gezeigt, daß die Anwendungen von Regel I, II, III in den Zeilen der Tabelle in dieser Reihenfolge auftreten. Formulieren Sie eine Vermutung über die Reihenfolge des Auftretens in den Spalten der Tabelle! Läßt sich über Regel II noch mehr sagen? (2 Punkte)

Aufgabe 6.2 In der Vorlesung wurden die Begriffe *maximales Paar* (i, j, k) und *maximaler Repeat* (*wiederholter Teilstring*) eingeführt und Eigenschaften dazu bewiesen. Formulieren Sie eine Hypothese über die maximale Anzahl von maximalen Repeats in einem String der Länge n und beweisen Sie sie (2 Punkte).

Aufgabe 6.3 Für die Implementierung eines Suffixbaums für einen String muß man sich eine geeignete Datenstruktur für die inneren Knoten überlegen. Ein Kriterium dafür ist, wie die vom Knoten ausgehenden Kanten im Knoten repräsentiert werden. Dabei spielen sowohl der benötigte Speicherbedarf als auch die Zugriffszeit auf eine gesuchte Kante (um sie zu verfolgen) eine wichtige Rolle.

Wenn Σ das Alphabet aller möglichen Zeichen im String ist, dann gibt es maximal $|\Sigma|$ ausgehende Kanten pro inneren Knoten. Damit können alle Kanten mit einem Feld der Länge $|\Sigma|$ repräsentiert werden (der erste Buchstabe der Kantenbeschriftung wird als Index verwendet). Alternativ können ausgehende Kanten über eine verkettete Liste oder einen balancierten binären Suchbaum im Knoten repräsentiert werden.

Vergleichen Sie die drei Möglichkeiten hinsichtlich des Zeitbedarfs zum Finden einer Kante und des Speicherbedarfs (ein Verweis auf ein Listenelement bzw. einen Kindknoten benötige genausoviel Speicher wie ein Verweis auf eine Kante). Überlegen Sie sich, in welchen Ebenen des Suffixbaums (obere, mittlere, untere) Sie welche Art der Repräsentation anwenden würden und begründen Sie Ihre Entscheidung. (3 Punkte)

Abgabe per E-Mail an grau@informatik.uni-halle.de oder in Raum 4.12.