Prof. Dr. Gerhard Reinelt
Dipl.-Math. Hanna Seitz
Institut für Informatik
Universität Heidelberg

Algorithmen & Datenstrukturen

12. Übungsblatt SS 07
Abgabetermin: 18.07.2007

Aufgabe 43

Erläutern Sie, wie man die Vorkommen eines Suchwortes P in einem Text T finden kann, indem man die Präfixfunktion π des Strings PT (der String der Läenge $m+n$, bei dem es sich um die Verknüpfung von P und T handelt) untersucht.

Aufgabe 44

Konstruieren Sie einen endlichen Automaten zur Erkennung des Musters $P =$`abacabab` und illustrieren Sie dessen Funktionsweise auf der Zeichenkette $T =$`ababacbabacababacabab`.

Aufgabe 45

Führen Sie die Schritte des Algorithmus von Knuth, Morris und Pratt für den Eingabetext

`aabcaabcaabcac`

und das Suchwort

`abcaabc`

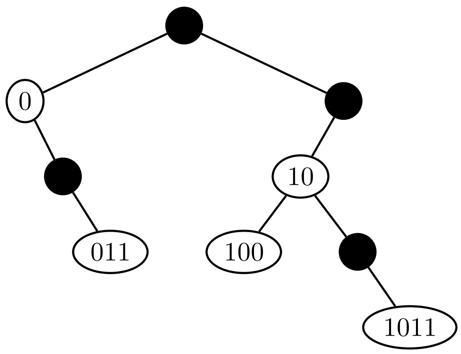
durch.

Aufgabe 46

Gegeben seien zwei Zeichenketten $a = a_0a_1 \dots a_p$ und $b = b_0b_1 \dots b_q$, wobei a_i und b_j einer geordneten Menge von Zeichen entnommen sind. Die Zeichenkette a heißt **lexikographisch kleiner** als die Zeichenkette b, falls entweder

- ein Index j, $0 \le j \le \min(p, q)$, existiert, so dass $a_i = b_i$ für alle $i = 0, 1, \dots, j-1$ und $a_j < b_j$, oder
- $p < q$ und $a_i = b_i$ für alle $i = 0, 1, \dots, p$.

Die unten abgebildete Datenstruktur **Radix Baum** speichert die Bit-Zeichenketten 1011, 10, 011, 100 und 0. Wenn nach einem Schlüssel $a = a_0a_1 \dots a_p$ gesucht wird, verzweigt man an einem Knoten in Tiefe i nach links, wenn $a_i = 0$ und nach rechts, wenn $a_i = 1$. Knoten, deren zugehörige Schlüssel nicht im Baum enthalten sind, sind schwarz gefüllt dargestellt.



Sei S eine Menge verschiedener binärer Zeichenketten, deren Längen aufsummiert n ergibt. Zeigen Sie, wie sich ein Radix Baum verwenden läßt, um S lexikographisch in $\Theta(n)$ Zeit zu sortieren. Die Ausgabe für den obigen Baum sollte 0, 011, 10, 100, 1011 sein.