Technische Universität München
Fakultät für Informatik
Lehrstuhl für Effiziente Algorithmen
Dr. Hanjo Täubig
Johannes Nowak

Sommersemester 2008
Übungsblatt 1
22. April 2008

Algorithmische Bioinformatik I

Aufgabe 1

Bestimmen Sie für die Zeichenreihe

01011010100

die Tabelle $border[]$ für den Knuth-Morris-Pratt-Algorithmus.

Aufgabe 2

Modifizieren Sie den KMP-Algorithmus so, daß er für jedes Vorkommen einer Zeichenreihe s in t die Position des Vorkommens ausgibt.

Aufgabe 3

Geben Sie eine Optimierung des KMP-Algorithmus für den Fall an, dass das vorliegende Alphabet binär ist. Definieren Sie hierzu die $border[]$-Tabelle geeignet um und modifizieren Sie den KMP-Algorithmus. Wie sieht die Tabelle für das Beispiel in Aufgabe 1 aus?

Aufgabe 4

a) Bei der Anwendung eines String Matching Algorithmus auf einen Text $T = t_0 t_1 \cdots t_{n-1}$ und ein Suchmuster $S = s_0 s_1 \cdots s_{m-1}$ $(m < n)$ trete folgendes Szenario auf: Der Algorithmus untersucht durch vollständiges Scannen von links nach rechts, ob S in T ab Position i enthalten ist und stellt das erste Mismatch an Position j in S fest, d.h. $t_{i+j} \neq s_j$, wobei $j > 0$. Daraufhin wird S um eine Position nach rechts geschoben und festgestellt, daß hier ein vollständiges Vorkommen von S in T vorliegt. Welche Form hat der String S?

b) Wie sieht S aus, wenn das erste vollständige Vorkommen von S in T erst an Position $i + 2$ oder $i + 3$ vorliegt?