

Übung zu Algorithmen auf Sequenzen Blatt 3

Ausgabe: 25.10.2018 **Besprechung:** 15.11.2018

Sei $P = \text{AGTACGAG}$.

Sei $T = \text{AACGTAAGTACGAGAGTACG}$.

Aufgabe 3.1

Erstellen Sie den NFA für das Muster P . Erstellen Sie für den Shift-And-Algorithmus die entsprechenden Masken für P über dem Alphabet $\Sigma = \{\text{A, C, G, T}\}$. Führen Sie eine Mustersuche mit dem Shift-And auf dem Text T aus.

Aufgabe 3.2

Kehren Sie die Logik des Shift-And-Algorithmus um (0 bedeutet aktiver Zustand, 1 bedeutet inaktiver Zustand), um den Shift-Or-Algorithmus zu erstellen. Welchen Vorteil hat das?

Aufgabe 3.3

Erstellen Sie den DFA für das Muster P , und stellen Sie seine lps-Funktion auf. Führen Sie den Knuth-Morris-Pratt-Algorithmus auf dem Text T aus.

Aufgabe 3.4

Erstellen Sie die Sprungtabelle für den Horspool-Algorithmus für das Muster P über dem Alphabet $\Sigma = \{\text{A, C, G, T}\}$. Bestimmen Sie die Anzahl und die Länge der Sprünge, die bei der Mustersuche für den Text T ausgeführt werden.

Aufgabe 3.5

Was sind Gründe für die schlechte worst-case Laufzeit des Horspool-Algorithmus? Unter welchen Umständen tritt sie auf?

Aufgabe 3.6

Bestimmen Sie die Anzahl und die Länge der Sprünge, die bei der Mustersuche mittels BNDM-Algorithmus für das Muster P auf dem Text T ausgeführt werden.

Zusatzaufgabe 3.7

Sei $A(q)$ die eindeutige aktive Zustandsmenge im NFA mit Maximum q . Für q mit $|A(q)| \geq 2$ sei $p(q)$ das zweitgrößte Element in $A(q)$. Zeigen Sie, dass dann $A(p(q)) = A(q) \setminus \{q\}$ gilt.

Zusatzaufgabe 3.8

Schreiben Sie explizit die zwei Phasen des Knuth-Morris-Pratt-Algorithmus (Vorbereitung + Textsuche) mit der p -Funktion aus der vorigen Aufgabe statt mit der lps -Funktion auf.

Zusatzaufgabe 3.9

Überlegen Sie sich eine Erweiterung des Horspool-Algorithmus, so dass Sprünge ggf. auf den letzten beiden Zeichen des aktuellen Textfensters basieren. Dadurch sollten die Sprünge häufiger länger werden. Denken Sie, dass der Algorithmus dadurch insgesamt schneller wird? Testen Sie Ihre Hypothese an langen(!) zufälligen Texten und Mustern bei verschiedenen Musterlängen und Alphabetgrößen.