

Übungen zur Vorlesung Algorithmen auf Sequenzen

TU Dortmund, SS 2008

Prof. Dr. Sven Rahmann

Blatt 1 vom 07.04..2008

Abgabe am Fr 11.04.2008 in der Vorlesung

Aufgabe 1 Wie viele *verschiedene* Teilstrings kann ein String der Länge n minimal / maximal enthalten (bei beliebigem Alphabet)? Gib jeweils einen minimierenden / maximierenden String an. Beantworte dieselbe Frage für Präfixe, Suffixe, und Teilsequenzen.

Aufgabe 2 Ein *verallgemeinerter String* (*generalized string*) über A ist ein String über der Potenzmenge von A , wobei an keiner Stelle die leere Menge zugelassen ist. Ein verallgemeinerter String über A kann auf natürliche Weise als eine Menge von Strings über A interpretiert werden. Bei DNA-Sequenzen (Sequenzen über Nukleinsäuren) hat sich hierzu der IUPAC-Code durchgesetzt, der die nichtleeren Teilmengen des DNA-Alphabets $\{A,C,G,T\}$ durch neue Symbole beschreibt; siehe <http://www.dna.affrc.go.jp/misc/MPsrch/InfoIUPAC.html>.

- Informiere dich über den molekularen Aufbau von DNA; wofür stehen A, C, G, T?
- Gib die Menge der DNA-Strings an, die durch den IUPAC-String CRASH beschrieben werden.
- Lerne den IUPAC-Code fuer Nukleinsäuren auswendig.

Aufgabe 3 Zeichne den DFA für den String ababaca.

Aufgabe 4 Beweise die in der Vorlesung aufgestellte Behauptung

$$\text{state}(x) = \text{lps}(x) \quad \text{für alle Strings } x.$$