

Übung zu Algorithmen auf Sequenzen Blatt 1

Ausgabe: 16. Oktober 2014 **Besprechung:** 30. Oktober 2014

Aufgabe 1.1

Es gibt 2^4 binäre Strings der Länge $n = 4$. Gib einen möglichst kurzen String an, in dem alle 16 verschiedenen Strings als Substrings vorkommen. Wie ist die Minimallänge eines solchen Strings? Gib zur Berechnung der Länge eine Formel an.

Aufgabe 1.2

Wie viele Teilstrings und wie viele Teilsequenzen hat eine Sequenz der Länge n stets? Wie viele verschiedene Teilstrings / Teilsequenzen eine Sequenz der Länge n hat, hängt allerdings von der Sequenz selbst ab. Gib hierfür untere und obere Schranken an, sowie die Extrembeispiele.

Aufgabe 1.3

Wie viele Bit-Sequenzen der Länge n mit k Einsen gibt es? Wie viele sind das konkret für $n = 20$ und $k = 10$? Welcher Anteil aller Bit-Sequenzen der Länge 20 ist das?

Aufgabe 1.4

Sei $\Sigma = \{0, \dots, K-1\}$. Seien $a, b \in \Sigma$ zwei Zeichen, die unabhängig voneinander zufällig gewählt werden. Wie groß ist die Wahrscheinlichkeit, dass $a = b$ ist, wenn alle Zeichen gleich wahrscheinlich sind? Wie groß ist die Wahrscheinlichkeit, dass $a = b$ ist, wenn Symbol k die Wahrscheinlichkeit p_k hat mit $p_k \geq 0$ für alle $k = 0, \dots, K-1$ und $\sum_k p_k = 1$?

Aufgabe 1.5

Wie viele zusätzliche Bits benötigt man genau, um die in der Vorlesung genannte rank-Datenstruktur zu erstellen, wenn die Eingabesequenz eine Länge von $n = 2^{16} = 65536$ hat?

Wie sieht diese Datenstruktur konkret für die Sequenz $s = 010101 \dots$ der Länge 65536 aus?