

Aufgabe 4.1.

Ein String ω ist ein *prefix tandem repeat* des Strings S wenn gilt:

- ω ist ein Präfix von S
- es gibt einen String u mit $uu = \omega$

Geben Sie einen linearen Algorithmus an, welcher das längste *prefix tandem repeat* von S findet.

Definition: Ein maximales Repeat, das nicht echter Teilstring eines anderen Repeats ist, heißt supermaximales Repeat.

Lemma: Ein String ω ist ein supermaximales Repeat gdw.

- das ω -Intervall $[i..j]$ ist ein lcp-Intervall mit lcp-Wert $\ell = |\omega|$ für das gilt: $LCP[k] = \ell$ für alle k mit $i + 1 \leq k \leq j$,
- die Buchstaben $BWT[i], BWT[i + 1], \dots, BWT[j]$ sind paarweise verschieden.

Aufgabe 4.2.

Sei ω ein längstes Repeat. Zeigen Sie, dass ω ein supermaximales Repeat (und damit auch ein maximales Repeat) sein muss.

Aufgabe 4.3.

Entwerfen Sie einen linearen Algorithmus welcher den lcp-Intervallbaum Bottom-Up durchläuft und dabei simultan alle maximalen und supermaximalen Repeats ausgibt.

Aufgabe 4.4.

Zeichnen Sie einen de Bruijn Graph für $k = 2$ über dem Alphabet $\Sigma = \{a, c, g, t\}$. Finden Sie einen Eulerkreis in dem Graph und geben Sie die korrespondierende nicht-zyklische de Bruijn Sequenz an. Wie viele supermaximale Repeats enthält die Sequenz?