

Algorithmen zur Sequenzanalyse

Wintersemester 2019/2020
Besprechung am 17.12.2019

Übungsblatt 5

Prof. Dr. E. Ohlebusch,
Institut für Theoretische Informatik

Aufgabe 5.1.

Zeichnen Sie den 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?

Aufgabe 5.2.

Der String $mmnm\$uululi_ _ _$ ist BWT eines Strings S , wobei das Zeichen $_$ als Ersatz für ein Leerzeichen verwendet wird. Rekonstruieren Sie S .

Aufgabe 5.3.

Sei S ein String der Länge n ohne Abschlusszeichen $\$$ und L die letzte Spalte der Matrix, die alle n Rotationen von S in lexikografischer Reihenfolge enthält. Sei SA das Suffix Array von S und der String $BWT[1..n]$ definiert durch:

$$BWT[i] = \begin{cases} S[SA[i] - 1] & \text{falls } SA[i] > 1 \\ S[n] & \text{falls } SA[i] = 1 \end{cases}$$

Finden Sie einen String S für den $L \neq BWT$ gilt.

Aufgabe 5.4.

Erweitern Sie Algorithmus 15 des Skripts so, dass zusätzlich das Suffix Array des Strings berechnet wird. Ist es möglich, das LF-Array mit dem Suffix Array zu überschreiben um Speicherplatz zu sparen?