

Algorithmen zur Sequenzanalyse: Übungsblatt 7

Enno Ohlebusch
Timo Beller

Das Übungsblatt wird am 04.02.2014 besprochen.

Aufgabe 1

Die Burrows-Wheeler-Transformationen der Strings S und S^{rev} sind jeweils in einem Wavelet-Baum gespeichert, siehe Abbildung 1 und 2. Ermitteln Sie mittels Algorithmus 29 im Skript, wie oft der String GTA in S vorkommt.

Aufgabe 2

Erweitern Sie die Vorkommen (beginnend mit den gefundenen Vorkommen aus Aufgabe 1) jeweils mit einem Rückwärtssuchschritt um den Buchstaben c und versuchen Sie mit einem Vorwärtssuchschritt, das Muster auf der anderen Seite mit dem Watson-Crick-Komplement von c zu erweitern. Wie lautet das längste Muster, das Sie auf diese Weise finden können? Die Watson-Crick-Paarungen sind $A-T$ und $C-G$.

Aufgabe 3

Zeigen Sie, dass die links-rechts Zerlegung von P bzgl. S eindeutig ist.

Aufgabe 4

Geben Sie einen Algorithmus an, der mit Hilfe der BWT von S^{rev} das M_{lr} -Array berechnet. Wie ist die Laufzeit Ihres Algorithmus? Wie kann das M_{lr} -Array effizient gespeichert werden?

Aufgabe 5

Berechnen Sie mit Hilfe von Algorithmus 27 im Skript das LCP-Array für den String $S = annasanannas\$. Die Burrows-Wheeler-Transformierte von S ist $BWT = ssn\$nnannaaaa.$$

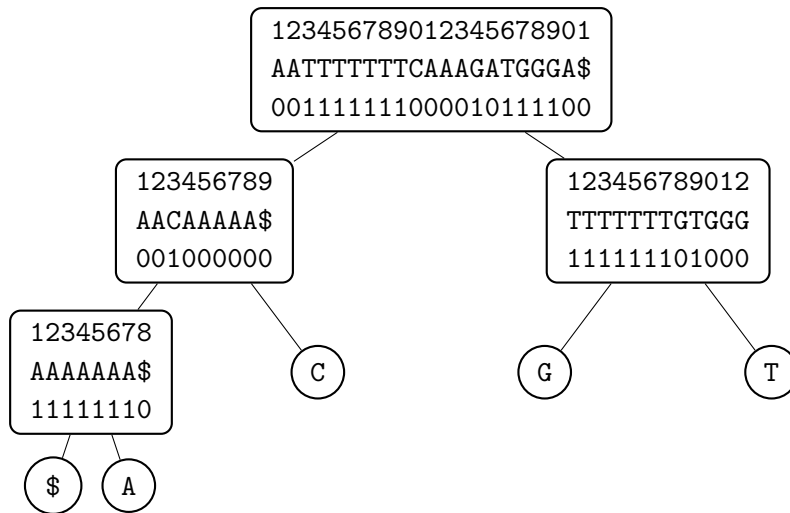


Abbildung 1: Wavelet-Baum von *AATTTTTTTCAAAGATGGGA\$*

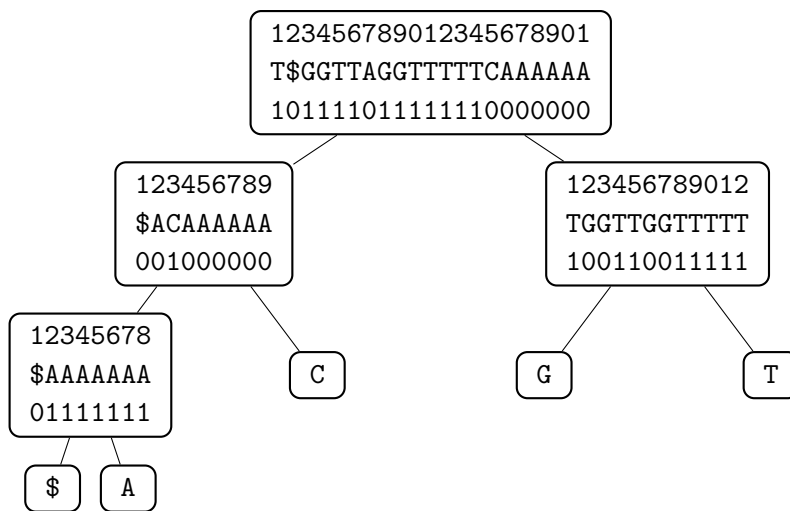


Abbildung 2: Wavelet-Baum von *T\$GGTTAGGTTTTTCAAAAAA*