

Algorithmen auf Sequenzen

Übungsblatt 12

Abgabetermin: Dienstag, 31.01.2017, 10 Uhr
(Via email an sophie.friedl@bio.ifi.lmu.de oder persönlich)

Hinweis: Für die Teilnahme an der Klausur am 8.2.2017 ist eine Anmeldung (bis Sonntag 29.1.2017) über https://www.bio.ifi.lmu.de/studium/ws2016/vlg_sequ/klausuranmeldung/index.html nötig. Bioinformatik-Studenten müssen sich zusätzlich auch über TUMOnline anmelden.

1. Aufgabe (Burrows-Wheeler-Transformation):

Betrachten Sie das Wort $t = t_1 \cdots t_{17} = \text{ABANANAISANANANAS}$ und die zugehörige Burrows-Wheeler-Transformierte $\hat{t} = \hat{t}_1 \cdots \hat{t}_{18} = \text{S\$NNBSNNNAAAAAAAAAI}$.

- Konstruieren Sie die für die Rank-Select-Datenstruktur benötigten Arrays R , R' und R'' jeweils für B_A und B_N (für Anfragen vom Typ $\text{rank}_0(\cdot)$). Hierbei soll $s' = 3$ und $s = 6$ verwendet werden, auch wenn dann $(s')^2 \neq s$.
- Beantworten Sie die Rank-Anfragen $\text{rank}_0^{B_A}(17)$ und $\text{rank}_1^{B_N}(17)$ nach der in der Vorlesung vorgestellten Methode basierend auf den Arrays aus (a).
- Konstruieren Sie den Wavelet-Tree zu \hat{t} für $t\$$.
- Bestimmen Sie die Werte von $\text{Occ}(A, 14)$ und $\text{Occ}(N, 14)$ nach der in der Vorlesung vorgestellten Methode aus dem Wavelet-Tree aus (c).

Hinweis: Das zu betrachtende Alphabet ist $\Sigma \cup \{\$\} = \{\$, A, B, I, N, S\}$, wobei die Ordnung auf dem Alphabet durch die Reihenfolge gegeben ist.

2. Aufgabe (Permutationen):

Finden Sie eine Permutation auf 4 Elementen, bei deren optimaler Sortierung durch (ungerichtete) Reversionen ein Strip aufgebrochen werden muss.

3. Aufgabe (Min-SBR):

Wenden Sie den Algorithmus zur 2-Approximation zur Bestimmung der minimalen Reversal-Distanz auf $(6, 7, 3, 4, 5, 8, 9, 1, 2)$ an. Geben Sie dabei alle Zwischenschritte an und erläutern Sie, warum eine bestimmte Reversion angewendet wird.