Übungen zur Vorlesung Sequenzanalyse

 Universität Bielefeld, SS 2022 Prof. Dr. Jens Stoye · Dr. Marília D. V. Braga

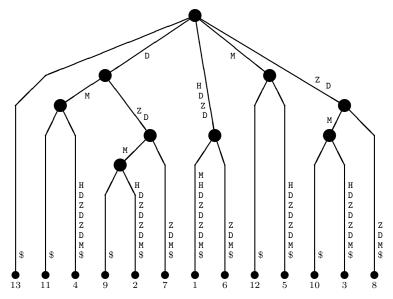
https://gi.cebitec.uni-bielefeld.de/teaching/2022summer/sa

Übungsblatt 9 vom 2.6.2022 Abgabe am 9.6.2022 bis 12:00 Uhr (mittags)

Aufgabe 1 (Suffixbaum und lcp-Array)

(3 Punkte)

Gegeben seien die Sequenz $s = \mathtt{HDZDMHDZDZDM}$ und der Suffixbaum von s\$ mit lexikographisch sortierten Kanten:



Berechne mit Hilfe dieses Baums das lcp-Array von s, ohne rank oder pos explizit auszurechnen.

Aufgabe 2 (Manber-Myers-Algorithmus)

(14 Punkte)

Gegeben sei die Sequenz $s={\tt DBADACDBCBADBADADAC}.$

- 1. Erstelle das Suffix-Array pos(s\$) unter Verwendung des Manber-Myers-Algorithmus. Gib das Zwischenergebnis nach jeder Phase des Algorithmus' an.
- 2. Berechne die Arrays rank(s\$) und lcp(s\$).
- 3. Verwende Binärsuche im Suffix-Array pos(s\$), um alle Vorkommen des Musters $p=\mathtt{ADA}$ in s zu finden. Gib alle Zwischenschritte an.

Aufgabe 3 (Manber-Myers-Algorithmus)

(3 Punkte)

Wie viele Phasen benötigt (im $worst\ case$) der Manber-Myers-Algorithmus für einen String der Länge n? Gib eine Beispielsequenz an, bei der diese Anzahl Phasen erreicht wird.