

Präsenzübungen 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>

Präsenzübungsblatt 9, vom 7.6.2022

Aufgabe 1 (Tries)

Ein Suffixtrie ist ein Σ -Baum aller Suffixe einer Sequenz. Zeige, dass die Anzahl der Knoten eines Suffixtrie einer Sequenz der Länge n in $O(n^2)$ liegt.

Aufgabe 2 (Suffixbaum-Suffixarray)

Gegeben ist der String $s = \text{CATATAGTAGCGTCATAGT}$.

1. Zeichne den Suffixbaum für $s\$$, sortiere dabei alle ausgehenden Kanten lexikographisch (wobei $\$ < A < C < G < T$).
2. Beschrifte die Blätter mit dem Start-Index des zugehörigen Suffixes in $s\$$. Die Indizierung beginnt bei 0.
3. Berechne mit Hilfe dieses Baums das lcp-Array und das Suffix-Array pos von $s\$$.
4. Verwende Binärsuche im Suffix-Array $\text{pos}(s\$)$, um alle Vorkommen der folgenden *Patterns* in s zu finden und gib ihre Startpositionen an:
 - $p_1 = \text{CAT}$
 - $p_2 = \text{TAG}$
 - $p_3 = \text{ATA}$