

Präsenzübungen zur Vorlesung Sequenzanalyse

Universität Bielefeld, SS 2019

Dr. Daniel Dörr · Michel T. Henrichs

<https://gi.cebitec.uni-bielefeld.de/teaching/2019summer/sa>

Präsenzübungsblatt 7, vom 27.5.2019

Aufgabe 1 (Suffixbäume)

Gegeben ist der String $s\$ = \text{TACGCACGGCAAGCCCGTGC\$}$, wobei $\$ < A < C < G < T$.

1. Zeichne den Suffixbaum für s , sortiere dabei alle ausgehenden Kanten lexikographisch.
2. Beschrifte die Blätter mit dem Start-Index des zugehörigen Suffixes in s . Die Indizierung beginne bei 0.
3. Beschrifte jeden Knoten mit der Anzahl der unter ihm liegenden Blätter.
4. Suche die folgenden 10 Pattern im Suffixbaum und gib ihre Startpositionen an:

- GCA
- CGTG
- GCAC
- TGC
- CGCAC
- GC
- AAGC
- CGCA
- ACGC
- CA

5. Erzeuge das Suffixarray zu $s\$$.