Präsenzübungen zur Vorlesung Sequenzanalyse

Universität Bielefeld, WS 2016/2017
Prof. Dr. Jens Stoye · M.Sc. Tizian Schulz

https://gi.cebitec.uni-bielefeld.de/teaching/2016winter/sa

Präsenzübungsblatt 5, Woche 48/2016

Aufgabe 1 (Suffixbäume)
Gegeben ist der String $s\$ =$ `ATAATTATTTCGCCGGCGGG$`, 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 beginnt bei 1.
3. Beschrifte jeden Knoten mit der Anzahl der unter ihm liegenden Blätter.

Aufgabe 2 (Suche im Suffixbaum)
Suche die folgenden 10 *Pattern* im Suffixbaum aus Aufgabe 1:

- `CGGG`
- `TTA`
- `TAC`
- `AATA`
- `TTCGC`
- `AATTA`
- `GCCGC`
- `GGCG`
- `TAATT`
- `ATTAC`

Aufgabe 3 ($ – Der Wächter)
Bei der Erstellung eines Suffixbaums wird am Ende des Strings ein eindeutiges Zeichen, z.B. $\$ \in \Sigma$, angehängt. Warum?