

Präsenzübungen zur Vorlesung Sequenzanalyse II

Universität Bielefeld, SS 2011
Prof. Dr. Jens Stoye · Daniel Dörr

<http://wiki.techfak.uni-bielefeld.de/gi/Teaching/2011summer/SequenzAnalyse>

Präsenzübungsblatt 2

Aufgabe 1 (Knoten und Kanten im Suffixbaum) Sei v innerer Knoten mit $\text{string}(v) = ax$ mit $a \in \Sigma$ und $x \in \Sigma^*$; d.h. $\text{node}(ax)$ existiert und ax ist rechtsverzweigend. Zeige, dass dann auch $\text{node}(x)$ existiert.

Bemerkung: Dies erlaubt es, stets sogenannte Suffixlinks zu definieren, die von einem Knoten $\text{node}(ax)$ auf den Suffixknoten $\text{node}(x)$ zeigen. Es ist ja nicht selbstverständlich, dass so ein Knoten existiert!

Aufgabe 2 (Shortest Unique Substring Problem)

Gebe einen Algorithmus im Pseudocode an, welcher unter Verwendung eines Suffixbaumes das Shortest Unique Substring Problem löst. Mache dir dabei folgende Eigenschaften des Suffixbaumes $s\$$ zunutze:

- Wenn ein String w mindestens zweimal in s vorkommt, dann gibt es mindestens zwei Suffixes in $s\$$ von denen w ein echtes Prefix ist. Daher entspricht w in einem Suffixbaum über $s\$$ einem Pfad, welcher in einem Knoten mit verzweigenden Kanten endet.
- Wenn ein String w nur einmal in s vorkommt, dann gibt es nur ein Suffix in $s\$$ von dem w ein Präfix ist. Daher entspricht w in einem Suffixbaum über $s\$$ einem Pfad, welcher in einem Blattknoten endet.