

Übungen zur Vorlesung Sequenzanalyse II

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

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

Blatt 1 vom 29.04.2011

Abgabe am 6. Mai zu Beginn der Vorlesung
Bitte gib auch den Namen deiner Tutorin an.

Aufgabe 1 (Suffixbäume)

(4 Punkte)

- Beschreibe zwei Anwendungen aus der bioinformatischen Praxis, für die man einen Suffixbaum verwenden kann.
- Gegeben ist der String $s\$ = \text{abcacbbaca\$}$, wobei $\$ < a < b < c$.
 - Zeichne den Suffixbaum für s , sortiere dabei die Blätter lexikographisch.
 - Beschrifte die Blätter mit dem Start-Index des zugehörigen Suffixes in s . Die Indizierung beginnt bei 0.
 - Beschrifte jeden Knoten mit der Anzahl der unter ihm liegenden Blätter.
 - Gibt es einen *längsten* maximalen Repeat in s ? Wie kann man diesen mit Hilfe des Suffixbaums finden?
- Zeige, dass der Speicherverbrauch eines Suffixbaums linear bezüglich der Eingabe ist.

(4 Punkte)

Aufgabe 2 (Laufzeit- und Speichereffizienz)

In den inneren Knoten eines Suffixbaumes müssen auf irgendeine Weise die Kinder gespeichert werden. Wie viel Platz insgesamt (bei Textlänge n und Alphabetgröße σ) und wie viel Zeit zum Auffinden einer bestimmten ausgehenden Kante benötigt man jeweils bei Verwendung

- eines Arrays fester Größe σ ?
- eines sortierten Arrays aller Kinder?
- einer verketteten Liste aller Kinder?
- eines balancierten Suchbaumes aller Kinder?
- einer Hash-Tabelle aller Kinder?

(4 Punkte)

Aufgabe 3 (Effiziente Speicherung eines Suffixbaums)

- Lies im Skript Abschnitt 8.4 über die effiziente Speicherung von Suffixbäumen.
- Ein Suffixbaum kann effizient als Array gespeichert werden, wie es auch in dem aufgeführten Beispiel von Abschnitt 8.4 gezeigt wird. Jeder Eintrag des Arrays entspricht dabei einem 32-Bit Integer-Wert, wobei 2 Bits zur Markierung verwendet werden.
Was ist die maximale Länge eines Strings (im *worst case*), welcher in der beschriebenen Implementierung gespeichert werden kann? (Tipp: Hilfreich ist Lemma 8.1 im Skript)