

## Foundations of Sequence Analysis

Winter 2005/2006

### Übungen

#### Beispielfragen zu Kapitel 4, Besprechung am 26.01.2006 bzw. 30.01.2006.

Die folgende Liste von Beispielfragen soll eine Hilfestellung bei der Wiederholung des Stoffs von Abschnitt 4 der Vorlesung „Grundlagen der Sequenzanalyse“ sein. Alle Fragen lassen sich mit Hilfe des Skripts *Foundations of Sequence Analysis* lösen. Die Liste definiert nicht den vollständigen Prüfungstoff zu der Vorlesung. Die Grundlage dafür ist das Skript.

1. Was ist ein  $\mathcal{A}^+$ -Baum, was ein kompakter  $\mathcal{A}^+$ -Baum?
2. Erläutern Sie die Begriffe *innerer Knoten*, *Blatt*, *verzweigender Knoten*, *innere Kante*, *Blattkante* und *Pfad* im Zusammenhang mit  $\mathcal{A}^+$ -Bäumen. Wie ist die Menge der in einem  $\mathcal{A}^+$ -Baum  $T$  vorkommenden Worte  $occ(T)$  definiert?
3. Wieso kann in einem  $\mathcal{A}^+$ -Baum jedem Knoten  $\alpha$  mit Pfad  $w = path(\alpha)$  eine eindeutige Bezeichnung  $\bar{w}$  zugeordnet werden?
4. Gegeben eine Sequenz  $S \in \mathcal{A}^+$ , wie ist der Suffixbaum von  $S$  definiert?
5. Welcher Zusammenhang besteht im Suffixbaum einer Sequenz  $S$  zwischen den Blättern und den Suffixen von  $S$ , welcher Zusammenhang besteht zwischen den verzweigenden Knoten und den rechtsverzweigenden Teilworten von  $S$ ?
6. Zeigen Sie, dass sich der Suffixbaum  $ST(S)$  in linearem Platz bezüglich der Länge von  $S$  abspeichern lässt. Skizzieren Sie eine konkrete Möglichkeit der Repräsentation.
7. Erläutern Sie den Algorithmus zur *Write-Only-Top-Down*-Konstruktion von Suffixbäumen. Was ist seine Laufzeit im schlechtesten Fall? Geben Sie ein Beispiel dafür an.
8. Was ist die grundlegende Strategie bei der *Online*-Konstruktion von Suffixbäumen nach dem Algorithmus von Ukkonen? Welches sind die entscheidenden Ideen, damit dieser in linearer Zeit läuft?
9. Was ist die grundlegende Strategie bei der *Online*-Konstruktion von Suffixbäumen nach dem Algorithmus von McCreight?
10. Erläutern Sie den Algorithmus zur exakten Textsuche mittels Suffixbäumen. In welchem Einsatzszenario ist dieser effizienter als Verfahren wie Boyer-Moore oder Knuth-Morris-Pratt?
11. Was sind minimale eindeutige Teilworte einer Sequenz und wie lassen sich diese effizient mit Hilfe von Suffixbäumen auffinden?
12. Was sind maximale eindeutige Matches zweier Sequenzen und wie lassen sich diese effizient mit Hilfe von (verallgemeinerten) Suffixbäumen auffinden?
13. Skizzieren Sie einen Algorithmus zum Auffinden von maximalen Repeats in einer gegebenen Sequenz. Was ist die Laufzeit des von Ihnen angegebenen Verfahrens?