

Übungen zur Vorlesung Grundlagen der Sequenzanalyse

Universität Bielefeld, WS 2006/07

Dr. Sven Rahmann · Dipl.-Bioinf. Katharina Jahn

<http://gi.cebitec.uni-bielefeld.de/teaching/2006winter/sequenzanalyse/>

Blatt 12 vom 18.01.2007

Abgabe am 25.01.2007 vor der Vorlesung um 8:30 in H3

Aufgabe 1 Gib einen Linearzeit-Algorithmus an, mit dem man einen längsten gemeinsamen Substring zweier gegebener Strings s und t finden kann, wenn das *Suffix-Array* von $s\#t$ gegeben ist. Dabei sind $\#$ und $\$$ zwei Zeichen, die im Alphabet nicht vorkommen.

Aufgabe 2 Gib einen Linearzeit-Algorithmus an, mit dem man einen kürzesten eindeutigen Substring eines Strings s finden kann, wenn der *Suffixbaum* von $s\$$ gegeben ist. Achtung: Die Substrings, die mit $\$$ enden, zählen hierbei nicht, sondern nur Substrings von s selbst.

Aufgabe 3 Gib einen String s_{\max} der Länge n an, so dass der Suffixbaum von $s_{\max}\$$ maximal viele innere Knoten hat. Gib auch einen String s_{\min} der Länge n an, so dass der Suffixbaum von $s_{\min}\$$ minimal viele innere Knoten hat. Das Alphabet darf dabei beliebig gewählt sein und von n abhängen.

Aufgabe 4 Formuliere eine angemessene Klausuraufgabe anhand des Vorlesungsstoffs und der bisherigen Übungsaufgaben.

Hinweise. Dies ist der letzte Zettel mit regulären Übungsaufgaben. Blatt 13 wird eine Art Übungsklausur (aber umfangreicher) sein. Die Klausur wird sich vor allem auf die Übungsaufgaben beziehen, aber natürlich auch voraussetzen, dass die Konzepte dahinter verstanden wurden.

Teilnahmevoraussetzungen zur Klausur sind 60% der Punkte und der A&D-Leistungsnachweis. Man kann die Klausur auch ohne diese Voraussetzungen zum persönlichen Vergnügen mitschreiben, jedoch werden wir die Klausur dann nicht korrigieren.

Zum Lernen sollten man, wo immer möglich, das neue Skript verwenden, und wenn dort noch Fragen offenbleiben, das alte Skript hinzuziehen.