technische universität dortmund

Bioinformatik / Lehrstuhl 11
Fakultät für Informatik

Sven Rahmann

Wintersemester 2014/2015

Übung zu Algorithmen auf Sequenzen
Blatt 3

Ausgabe: 13. November 2014 **Besprechung:** 27. November 2014

Aufgabe 3.1

Erläutere, wie der BNDM-Algorithmus auf dem Muster $P =$ `AGTACGAG` und dem Text $T =$ `AACGTAAGTACGAGAGTACG` arbeitet.

Aufgabe 3.2

Implementiere den BNDM-Algorithmus (in einer Sprache Deiner Wahl).

Teste ihn an 1000 zufälligen Mustern mit verschiedenen Alphabeten, indem du die gefundenen Positionen mit denen des naiven Algorithmus vergleichst.

Aufgabe 3.3

Welchen Algorithmus würde man in folgenden Szenarien wählen und warum?

1. T ist eine Abschlussarbeit (60 KB), P ein natürlichsprachliches Wort der Länge 7
2. T ist ein Bakteriengenom (5 Mbp), P eine DNA-Sequenz der Länge 25
3. T ist eine lange Bitsequenz (1 GBit), P eine Bitsequenz der Länge 256

Aufgabe 3.4 Erstelle für das erweiterte Muster `b-x(1,3)-a-b-x(1,1)-a` den entsprechenden NFA. Gib die Masken für `a`, `b`, sowie die Masken I und F an. Führe auf dem Text `bbabbabbaaababbabab` den erweiterten Shift-And-Algorithmus aus. Welche Zustände sind nach jedem Schritt jeweils aktiv?

Aufgabe 3.5

Konstruiere schrittweise den Suffixbaum zum Text `CAACACAAA$` mit Ukkonen's Algorithmus.

Zeige dann schrittweise, wie jeweils die Suche nach den Mustern `AACA` und `AACC` abläuft.

Aufgabe 3.6

Zeige ausführlich, dass ein Suffixbaum mit n Blättern maximal $n-1$ innere Knoten (inkl. Wurzel) und $2(n-1)$ Kanten hat.

Konstruiere eine Familie von Beispielen (für alle n), bei der die Maximalzahl erreicht wird.

Wie groß ist die minimale Anzahl an inneren Knoten und Kanten? Konstruiere auch hier eine Familie von Beispielen.