

Aufgabenblatt 12

Exklusionsmethode, BLAST, und FASTA

Abgabe: **01.02.2005** bis **17 Uhr** über Goya
Maximal: 20 Punkte
Namen nicht vergessen!

1. Gegeben ist das Pattern `GCCGTCCAGTTCG` und der Text `AATGCCGTGCAGTCTAGTTCCTCAATTCA`. Sucht mit der Methode von Baeza-Yates & Perleberger nach einem Match mit höchstens 3 Unterschieden. Skizziert dabei das Vorgehen schrittweise (Partitionierung, Suche, Check) und gebt am Ende die Position(en) an, wo das Pattern im Text gefunden wurde.

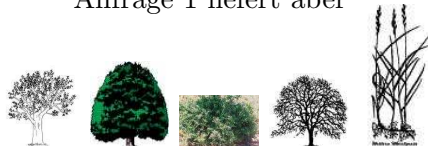
1 + 4 + 2 = 7 P

2. Wir haben eine Datenbank mit 10000 Bildern. Das Ergebnis, das wir auf unsere Anfrage erwarten sieht folgendermaßen aus:



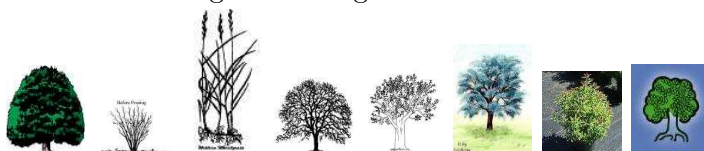
(a) Baum1 (b) Baum2 (c) Baum3 (d) Baum4 (e) Baum5 (f) Baum6

Anfrage 1 liefert aber



(g) Baum4 (h) Baum2 (i) Busch1 (j) Baum3 (k) Gras1

und Anfrage 2 das folgende:



(l) Baum2 (m) Busch3 (n) Gras1 (o) Baum3 (p) Baum4 (q) Baum6 (r) Busch2 (s) Baum5

Wie hoch sind die Sensitivität und Spezifität der beiden Anfragen jeweils? **2 P**

Wie ändern sich die beiden Werte, wenn wir nur noch die ersten 4 Ergebnisse zurückgeben?
1 P

3. In unserer Sequenzdatenbank soll nach folgendem Pattern mit der ursprünglichen BLAST-Methode gesucht werden:

P = GTTGCAT. Die Datenbank besteht nur aus dem String

S = ATTGGCCGTGTTGTATTTAGTTGCC.

Für die folgenden Aufgaben ist der Score für ein Match $m = 1$ und ein Mismatch bzw. Replace $r = -2$. Sucht mit der ursprünglichen BLAST-Methode, wobei $W = 4$ und $t = 3$ ist. Der geschätzte Schwellwert, c liegt bei 2.

6 P

4. Verwendet sowohl BLAST (<http://www.ncbi.nlm.nih.gov/BLAST/Blast.cgi>) als auch FASTA (<http://www.ebi.ac.uk/fasta33/>), um nach der Sequenz zu suchen, die auf der Webseite gegeben ist. Wie viele und welche der Sequenzen sind in beiden Ergebnismengen gleich?

2 P

Woran kann es liegen, dass die Ergebnismengen der beiden Methoden nicht komplett übereinstimmen?

2 P