Universität Bielefeld
Technische Fakultät

AG Genominformatik
Prof. Dr. Jens Stoye
Dipl.-Math. Julia Mixtacki
Dipl.-Inform. Arne Hauenschild
Dipl.-Inform. Wiebke Timm

Foundations of Sequence Analysis

Winter 2005/2006

Exercises

Übung 7, Besprechung am 19.12.2004 (14-16h in E0-160)
bzw. 15.12.2004 (14-16h in C01-148 und E01-108).

1. **Q-Gramm Modell.**

   Gegeben seien zwei Sequenzen **u** und **v**. Implementieren Sie ein Programm zur Berechnung der q-Gramm Distanz gemäß dem Algorithmus aus Abschnitt 3.6 des Skripts. Das Programm soll als weiteren Parameter den Wert von q haben.

2. **Fasta Ähnlichkeitsmodell.**

   Gegeben sei eine Query-Sequenz $w =$ `ATCACACTTAGC`. Die Datenbank besteht aus zwei Sequenzen $u_1 =$ `GCACACATC` und $u_2 =$ `ACTTAGT`. Berechnen Sie zunächst den Bucket $h(c)$ fuer w. Berechnen Sie dann den Fasta-Score für die Datenbank-Sequenzen.

3. **BlastP Ähnlichkeitsmodell.**

   Sei $\mathcal{A} =$\{C, G\} eine Teilmenge des DNA-Alphabets und σ eine Einheits-Score-Funktion

   $$\sigma(\alpha \to \beta) = \begin{cases} 1 & : \quad \alpha, \beta \in \mathcal{A} \wedge \alpha = \beta \\ 0 & : \quad sonst. \end{cases}$$

   (a) Gegeben sei eine Anfragesequenz $w = GGCCGC$. Konstruieren Sie (mit Papier und Stift) den DFA bezüglich dem BlastP Ähnlichkeitsmodell für $q = 4$ und $k = 3$.

   (b) Beschreiben Sie für ein beliebiges Alphabet in eigenen Worten:

   - Was ist der minimale Automat für $k = 0$?
   - Was ist der minimale Automat für $k > q$?
   - Was ist der minimale Automat für $k = 1$?
   - Was ist der minimale Automat für $k = q$?