Universität Bielefeld
Technische Fakultät

AG Genominformatik
Prof. Dr. Jens Stoye
Dipl.-Inform. Klaus-Bernd Schürmann

Foundations of Sequence Analysis

Winter semester 2003/2004

Exercises

Exercise 2, Discussion: 11/10/2003 and following dates.

1. Let u=AGTGCACACA and t= ATCACACTTA be two sequences. Calculate the dynamic programming matrices for the edit distance algorithm concerning different cost functions δ.

   (a) Unit cost:

   $$\delta(\alpha \to \beta) = \begin{cases} 0 & : \quad \alpha, \beta \in A \wedge \alpha = \beta \\ 1 & : \quad otherwise \end{cases}$$

   (b) Hamming cost:

   $$\delta(\alpha \to \beta) = \begin{cases} 0 & : \quad \alpha, \beta \in A \wedge \alpha = \beta \\ 2 & : \quad \alpha, \beta \in A \wedge \alpha \neq \beta \\ \infty & : \quad otherwise \end{cases}$$

   (c)

   $$\delta(\alpha \to \beta) = \begin{cases} 0 & : \quad \alpha, \beta \in A \wedge \alpha = \beta \\ 3 & : \quad \alpha, \beta \in A \wedge \alpha \neq \beta \\ 1 & : \quad otherwise \end{cases}$$

2. The alignment of two sequences mainly depends on the cost function.

   (a) What is the special property of the cost functions given in 1(b) and 1(c).

   (b) What properties should be fulfilled by a reasonable cost function.

3. Let δ be a cost function such that for each edit operation the following properties hold:

   $$\delta(\alpha \to \beta) \quad = \quad \delta(\beta \to \alpha) \qquad (symmetry) \qquad (1)$$
   $$\delta(\alpha \to \beta) = 0 \quad \Leftrightarrow \quad \alpha = \beta \qquad (zero\ property) \qquad (2)$$

   Show that for all $u, v \in \mathcal{A}^*$ the following properties hold:

   $$edist_\delta(u, v) \quad = \quad edist_\delta(v, u) \qquad (symmetry)$$
   $$edist_\delta(u, v) = 0 \quad \Leftrightarrow \quad u = v \qquad (zero\ property)$$

4. Implement the dynamic programming algorithm for the edit distance in C or Java.

5. Visualize the minimizing edges and paths of the edit graph concerning the strings u, t and the three cost functions given in 1. For each such graph write down all optimal alignments.