Universität Bielefeld
Technische Fakultät

AG Genominformatik
Dipl.-Inform. Klaus-Bernd Schürmann

Lecture: Spezielle Algorithmen der Sequenzanalyse
Summer semester 2006

Exercises

Exercise 1, Discussion: 04/12/2006.

1. **Edit distance.**
   (a) Given two arbitrary finite sequences u and v over a finite alphabet Σ and a cost function for edit operations $\delta : (\Sigma \cup \{\varepsilon\}) \times (\Sigma \cup \{\varepsilon\}) \to \mathbf{R}_0^+$, define the edit distance of u and v, $edist_\delta(u, v)$.

   (b) Given the two strings u = `GTCCA` and v = `GCCAA`. Calculate the dynamic programming matrices for the edit distance algorithm according to the unit cost function δ and visualize the edges of the minimizing path (could be more than one).

   Unit cost:

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

   (c) Calculate the edit distance of the sequences u = `ACCGG` and v = `ACCGCTGG` for unit costs.

   (d) Calculate all co-optimal alignments for the sequences u and v given in exercise 1(b).

   (e) Show that the edit distance is invariant with respect to string reversal:

   $$edist_\delta(u, v) \quad = \quad edist_\delta(u^{-1}, v^{-1})$$

   where w^{-1} is the reverse of string w, i.e. if $w = w_1 w_2 \dots w_k$ then $w^{-1} = w_k w_{k-1} \dots w_1$.

2. ***Q*-gram distance.**
   (a) Calculate the q-gram distance for the two sequence u = `TACTTTCTAGCTTA` und v = `ACTAGCTTTCTTAC`:
      i. for $q = 3$,
      ii. for $q = 5$.

   (b) Which of the two values for q can be used better to show that the q-gram distance is not a metric? Provide evidence for your answer.

3. ***O*-notation.**
   Rank the following functions by order of growth; that is, find an arrangement $g_1, g_2, \dots, g_{20}$ of the functions satisfying $g_1 = \Omega(g_2)$, $g_2 = \Omega(g_3), \dots, g_{19} = \Omega(g_{20})$.
   Partition your list into equivalence classes such that $f(n)$ and $g(n)$ are in the same class if and only if $f(n) = \Theta(g(n))$.
   $\log \log n$, $\sqrt{4}^{\log n}$, n^2, $n!$, $\left(\frac{3}{2}\right)^n$, n^3, $\log^2 n$, 2^{2^n}, $n \log n$, $n \cdot 2^n$, $2^{\log n}$, $\log n$, $\sqrt{n}$, $\log(n^2)$, $\sqrt{n}^4$, n^n, $2^3 \cdot n^3$, 4^n, 2^n, n^n