

Foundations of Sequence Analysis
Winter 2005/2006

Exercises

Übung 3, Besprechung am 14.11.2005 bzw. 17.11.2005.

1. Edit-Distanz — Eigenschaft.

Zeigen Sie, dass für beliebige Sequenzen \mathbf{u} und \mathbf{v} die Gleichung $edist_{\delta}(\mathbf{u}, \mathbf{v}) = edist_{\delta}(\mathbf{u}^{-1}, \mathbf{v}^{-1})$ gilt. Dabei ist \mathbf{u}^{-1} die Sequenz \mathbf{u} in umgekehrter Reihenfolge, d.h. falls $\mathbf{u} = u_1u_2 \dots u_m$, dann ist $\mathbf{u}^{-1} = u_mu_{m-1} \dots u_1$.

Zur Erinnerung einige Definitionen, die Sie für diese Aufgabe benötigen:

- Eine Kostenfunktion δ ordnet jeder Edit-Operation $\alpha \rightarrow \beta$ Kosten $\delta(\alpha \rightarrow \beta) \in \mathbf{R}_0^+$ zu, so daß gilt $\delta(\alpha \rightarrow \beta) = 0 \Leftrightarrow \alpha = \beta$.
- δ ist symmetrisch, falls $\delta(\alpha \rightarrow \beta) = \delta(\beta \rightarrow \alpha)$ für alle Edit-Operationen $\alpha \rightarrow \beta$.
- Die Kosten $\delta(A)$ eines Alignments A berechnen sich als die Summe der Kosten der Edit-Operationen in A .
- $edist_{\delta}(\mathbf{u}, \mathbf{v}) := \min\{\delta(A) \mid A \text{ ist ein Alignment von } u \text{ und } v\}$ ist die Edit-Distanz von \mathbf{u} und \mathbf{v} .

2. Edit-Distanz — Alignments.

Berechnen Sie für die folgenden Sequenzen die Edit-Distanz mit Hilfe des Editgraphen. Geben Sie drei optimale Alignments an:

$\mathbf{s}_1 = \text{GATTAC}$, $\mathbf{s}_2 = \text{GTACA}$.

- Bezüglich Einheitskosten.
- Bezüglich LCS-Kosten.