

Foundations of Sequence Analysis
Winter 2005/2006

Exercises

Übung 5, Besprechung am 28.11.2004 (14-16h in E0-160)
bzw. 01.12.2004 (14-16h in C01-148 und E01-108).

1. Edit-Graph – Dynamische Programmierung.

- (a) Definieren Sie den Edit-Graphen von zwei gegebenen Sequenzen \mathbf{u} und \mathbf{v} .
- (b) Erläutern Sie in eigenen Worten den Zusammenhang zwischen dem Edit-Graphen und dem dynamischen Programmieralgorithmus zur Berechnung der Edit-Distanz.

2. Einfache-Levenshtein-Distanz — Edit-Distanz.

Sei δ die Einfache-Levenshtein-Distanz. Beweisen Sie, dass folgende Gleichung für alle Strings \mathbf{u} und \mathbf{v} gilt:

$$2 \cdot lcs(\mathbf{u}, \mathbf{v}) + edist_{\delta}(\mathbf{u}, \mathbf{v}) = |\mathbf{u}| + |\mathbf{v}|$$

3. Einfache-Levenshtein-Distanz.

Berechnen Sie die LCS-Distanz von $\mathbf{u} = \text{CTATACT}$ und $\mathbf{v} = \text{ACATTGACG}$ (mit Stift und Papier) mit Hilfe des Greedy DP-Algorithmus.

- (a) Wie viele Matrixeinträge der Front-Matrix müssen berechnet werden?
- (b) Wie groß ist der Anteil der Matrix der nicht berechnet werden muss?

4. Sternchenaufgabe: Greedy-DP-Algorithmus.

Implementieren Sie den Greedy-DP-Algorithmus zur Berechnung der Einfachen-Levenshtein-Distanz.