# Übungen zur Vorlesung Sequenzanalyse

Universität Bielefeld, WS 2014/2015
Dr. Roland Wittler · Nina Luhmann · Linda Sundermann

`http://wiki.techfak.uni-bielefeld.de/gi/Teaching/2014winter/SequenzAnalyse`

**Übungsblatt 5 vom 11.11.2014**
**Abgabe in einer Woche vor Beginn der Vorlesung.**

**Aufgabe 1 (Approximatives Stringmatching)** (5 Punkte)

1. Finde die Endpositionen aller Vorkommen des Patterns $x =$ `TCTC` im Text $y =$ `CATCTCGACTC` mit maximal $k = 1$ Fehlern. Verwende die Cutoff-Variante von *Sellers'* Algorithmus mit Einheits**kosten** und markiere die *last essential indices.*
2. Gib zu jeder gefundenen Endposition alle zugehörigen Alignments an.
3. Warum macht es Sinn, in der Praxis nicht *alle* gefundenen Endpositionen auszugeben? (Stichwort: *Runs.*) Welche Endpositionen würde man im obigen Beispiel nicht ausgeben?

**Aufgabe 2 (Forward-Backward Technik)** (4 Punkte)

Gegeben sind die Sequenzen $s =$ `CACGTTAG` und $t =$ `GAGTCTA`. Berechne, jeweils mit Einheits**kosten**,

- die Edit-Matrix $D$,
- die Edit-Matrix $D^{\text{rev}}$ der reversen Strings,
- die Gesamtkostenmatrix $T$ und
- die Zusatzkostenmatrix $C$.

Gib ein optimales globales Alignment von $s$ und $t$ an und markiere es in $D$ und $C$.

**Aufgabe 3 (Paarweises Alignment mit linearem Speicherbedarf)** (4 Punkte)

Zeige anhand der Sequenzen aus Aufgabe 2, wie das optimale Alignment berechnet werden kann, während nur linearer Speicherplatz benötigt wird. Simuliere dazu die Schritte des in der Vorlesung besprochenen Divide-and-Conquer Verfahrens auf Kopien der Edit-Matrix. (Es kann teilweise auf die in Aufgabe 2 berechneten Werte zurückgegriffen werden.)

1. Markiere in einer neuen, leeren Matrix der entsprechenden Größe, für welche Bereiche $D$ und $D^{rev}$ berechnet werden würden bzw. berechne die Werte wenn nötig neu und trage sie ein.
2. Trage die Werte von $T$ für die mittlere Zeile $m' = \lceil m/2 \rceil$ ein.
3. Markiere ein Minimum in $T$ und die entsprechende Spalte, an der die Matrix geteilt wird.
4. Führe rekursiv die oberen Schritte für die beiden Teilmatrizen aus – bis jeweils nur noch eine Zeile übrig ist.