



## Blatt 13

**Aufgabe 13.1** In Aufgabe 12.4 b) haben wir festgestellt, dass für die Levensthein-Distanz und Strings  $S_1, S_2$  mit einer Distanz  $D(S_1, S_2) \leq k$  der Pfad des optimalen globalen Alignments in einem Schlauch der Breite  $k$  um die Hauptdiagonale verbleibt.

- Wir nehmen an, dass für zwei Strings  $S_1, S_2$  die Distanz  $D(S_1, S_2) \leq k$  ist. Entsprechend berechnen wir die Einträge in der DP-Matrix nur in einem Schlauch der Breite  $k$  um die Hauptdiagonale. Wie kann man aus der so gefüllten DP-Matrix schließen, ob unsere Annahme über  $D(S_1, S_2)$  korrekt war. (1 Punkt)
- Formulieren Sie einen Algorithmus, der die Ergebnisse aus Aufgabe a) nutzt, um die Suche nach einem optimalen globalen Alignment im average-case zu beschleunigen. (2 Punkte)
- Schätzen Sie die Laufzeit Ihres Algorithmus möglichst scharf nach oben ab. Vergleichen Sie die dann die worst-case Laufzeit Ihres Algorithmus mit der des Needleman-Wunsch Algorithmus. (2 Punkte)

**Aufgabe 13.2** Gegeben seien zwei Strings  $S$  und  $T$ .

**Definition:**

Eine *gemeinsame Teilsequenz* zweier Strings  $S$ ,  $|S| = m$ , und  $T$ ,  $|T| = n$ , ist eine Teilsequenz  $X$  mit  $X = S_{i_1}S_{i_2} \cdots S_{i_k} = T_{j_1}T_{j_2} \cdots T_{j_k}$  mit  $1 \leq i_1 < i_2 < \cdots < i_k \leq m$ ,  $1 \leq j_1 < j_2 < \cdots < j_k \leq n$ .

**Definition:**

Die *längste gemeinsame Teilsequenz* zweier Strings ist die gemeinsame Teilsequenz mit maximaler Länge.

Von nun an sollen nur noch Alignments betrachtet werden, die keine Mismatches enthalten.

- Warum reicht es für das Auffinden der längsten gemeinsamen Teilsequenz aus, nur die Operationen Match, Insert und Delete, nicht aber Mismatch/Replace zu erlauben? (1 Punkt)
- Gegeben sei ein Alignment  $A$  der Länge  $a$  ohne Mismatches. Wie lang ist eine daraus ableitbare gemeinsame Teilsequenz  $X$ ? Geben Sie eine Abschätzung der

Länge von  $X$  einerseits in Abhängigkeit von der Anzahl von Einfügungen und Löschungen, sowie andererseits in Abhängigkeit von den Längen  $m$  und  $n$  der beiden Strings an. (2 Punkte)

- b) Zeigen Sie, dass eine längste gemeinsame Teilsequenz aus einem Alignment mit minimaler Anzahl an Einfügungen und Löschungen resultiert. (1 Punkt)
- c) Nun soll ein globales optimales Alignment zweier Strings als Ausgangspunkt zur Bestimmung einer längsten gemeinsamen Teilsequenz durch dynamische Programmierung ermittelt werden. Wie ist die Kostenfunktion zu modifizieren, um sicherzustellen, dass das globale Alignment keine Mismatches enthält. (1 Punkt)

**Aufgabe 13.3** Arbeiten Sie Ihre Aufzeichnungen zur Vorlesung durch und stellen Sie eventuelle Unklarheiten fest. Formulieren Sie eine Fragestellung, die Sie in der nächsten Übung behandelt wissen möchten. (2 Punkte)

Abgabe per E-Mail an [grau@informatik.uni-halle.de](mailto:grau@informatik.uni-halle.de) oder in Raum 4.12.