

# Übungen zur Vorlesung Sequenzanalyse

Universität Bielefeld, SS 2021

Prof. Dr. Jens Stoye · Dr. Marília D. V. Braga

<https://gi.cebitec.uni-bielefeld.de/teaching/2021summer/sa>

## Übungsblatt 1 vom 22.4.2021

Abgabe am 29.4.2021 bis 12:00 Uhr (mittags)

### Aufgabe 1 (Komplexitätsklassen)

(4 Punkte)

Bei gegebener Eingabegröße  $n$  beschreibe  $f_1 = \frac{n^2}{4}$  die Anzahl Rechenschritte eines Algorithmus  $A_1$  und  $f_2 = 10n \cdot \log_{10} n$  die Anzahl Rechenschritte eines anderen Algorithmus  $A_2$ .

1. Stelle die Funktionen grafisch in einem gemeinsamen Koordinatensystem dar. Es genügt, wenn du die Funktionen im Bereich  $n = 0$  bis  $n = 5$  skizzierst.
2. Ermittle die Komplexitätsklassen der beiden Algorithmen in der  $\mathcal{O}$ -Notation.
3. Welcher Algorithmus hat die asymptotisch schlechtere Laufzeit? Ab welchem Wert für  $n$  wird das deutlich?

### Aufgabe 2 (Alphabete und Sequenzen)

(6 Punkte)

1. Gegeben  $k \geq 1$  und ein Alphabet  $\Sigma$ , wie ist  $\Sigma^k$  definiert? Wie viele Strings sind in  $\Sigma^k$ ? Wie sind  $\Sigma^*$  und  $\Sigma^+$  definiert und was ist der Unterschied zwischen ihnen?
2. Für das Alphabet  $\Sigma = \{\mathbf{A}, \mathbf{T}\}$ , schreibe alle Strings  $w \in \Sigma^3$  auf.
3. Gegeben sei der String  $s = \mathbf{ATGTGCTATATCGT}$ .
  - (a) Gib das (minimale) Alphabet von  $s$  an. Was ist die Länge von  $s$ ?
  - (b) Gib einen längsten Substring  $x$  von  $s$  an, so dass  $x \in \{\mathbf{A}, \mathbf{G}, \mathbf{T}\}^+$ .
  - (c) Gib die längste Subsequenz  $x$  von  $s$  an, so dass  $x \in \{\mathbf{A}, \mathbf{T}\}^+$ .
  - (d) Gib einen String an, der gleichzeitig echtes Präfix und echtes Suffix von  $s$  ist.

### Aufgabe 3 (Subsequenzen)

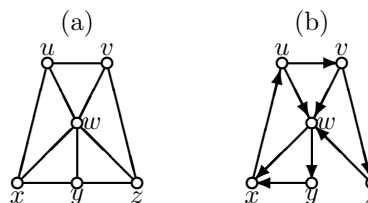
(2 Punkte)

Beschreibe einen Algorithmus, der für gegebene Strings  $x$  und  $y$  angibt, ob  $x$  Subsequenz von  $y$  ist. Welche Laufzeit hat der Algorithmus?

### Aufgabe 4 (Eulersche Graphen)

(2 Punkte)

Welche der folgenden Graphen haben einen Eulerpfad oder Eulerkreis? Warum? Wenn nicht, ist es möglich, durch Löschen einer Kante einen eulerschen Graphen zu erhalten?



### Aufgabe 5 (Edit-Sequenz und -Distanz)

(6 Punkte)

Überführe den String  $x = \mathbf{OKLAHOMA}$  in den String  $y = \mathbf{ALABAMA}$  durch Anwendung der *Edit*-Funktion.

1. Berechne die Edit-Sequenz  $E(x, \mathcal{DS}_A \mathcal{CCL}_B \mathcal{DS}_A \mathcal{DL}_M \mathcal{C})$ . Gib alle Zwischenschritte an. Wie hoch sind die Kosten der Edit-Sequenz?
2. Gib eine *Edit*-Sequenz mit minimalen Kosten an, die  $x$  in  $y$  überführt. Wie viele Edit-Sequenzen mit minimalen Kosten gibt es? Sind die *Edit*-Sequenzen mit minimalen Kosten immer die kürzesten? Begründe deine Antwort.
3. Gib eine *Edit*-Sequenz mit maximalen Kosten an, die  $x$  in  $y$  überführt. Wie viele Edit-Sequenzen mit maximalen Kosten gibt es? Sind die *Edit*-Sequenzen mit maximalen Kosten immer die längsten? Begründe deine Antwort.