

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 3 vom 6.5.2021

Abgabe am 13.5.2021 bis 12:00 Uhr (mittags)

Aufgabe 1 (O-Notation)

(2 Punkte)

Gib die O -Klassen der Funktionen $a(n)$ bis $e(n)$ an und sortiere sie anhand ihrer Untermengenrelation.

$$a(n) = n^2 \log_2(n) + 42$$

$$b(n) = 2^n + n^4$$

$$c(n) = 2^{2n}$$

$$d(n) = 2^n + 2^{n+3}$$

$$e(n) = \sqrt{n^5} + n^2$$

Hinweis: Die Aufgabe besteht darin, die O -Klassen zu identifizieren und sie zu verwenden, um die folgenden fünf Felder auszufüllen:

_____ \subset _____ \subset _____ \subset _____ \subset _____

Aufgabe 2 (Bäume und bipartite Graphen)

(3 Punkte)

Zeige, dass jeder Baum ein bipartiter Graph ist.

Aufgabe 3 (Maximal-Matches-Distanz)

(5 Punkte)

1. Gegeben seien die Sequenzen $x = \text{TATCTTCT}$ und $y = \text{AATCATATA}$.

(a) Berechne die folgenden Partitionen: $P_{\text{LR}}(x, y)$, $P_{\text{RL}}(x, y)$, $P_{\text{LR}}(y, x)$, $P_{\text{RL}}(y, x)$.

(b) Wie ist die Maximal-Matches-Distanz $\delta(x||y)$ von x bezüglich y definiert? Gib $\delta(x||y)$ und $\delta(y||x)$ an.

2. Zeige an einem von dir ausgedachten Beispiel, dass die Maximal-Matches-Distanz keine Metrik ist.

Aufgabe 4 (q -Gramm- und Maximal-Matches-Distanzen als Filter)

(7 Punkte)

Gegeben seien die Sequenzen:
$$\left\{ \begin{array}{l} x = \text{AATCGCGGTAC} \\ y_1 = \text{AATATCGGTGG} \\ y_2 = \text{AATCGAGGTAC} \\ y_3 = \text{AGGTACAATCG} \\ y_4 = \text{CTGAACGTCTG} \end{array} \right.$$

Wir wollen entscheiden, ob die Sequenzen y_1, \dots, y_4 eine Edit-Distanz von max. 1 zur Sequenz x haben können, ohne alle Edit-Distanzen zu berechnen.

1. Berechne die 2-Gramm-Profile der Worte x und y_1, \dots, y_4 . Filtere die Sequenzen y_1, \dots, y_4 mit Hilfe der 2-Gramm-Distanz. Welche Sequenzen können ausgeschlossen werden?

2. Filtere die übrigen Sequenzen mit Hilfe der Maximal-Matches-Distanz. Welche Sequenzen bleiben als Kandidaten übrig?

3. Nenne einen weiteren Filter, den man auch noch verwenden könnte.

Aufgabe 5 (Laufzeit der q -Gram-Distanz)

(3 Punkte)

Begründe, warum die Berechnung der q -Gram-Distanz von zwei Strings in linearer Zeit berechnet werden kann. Gehe dafür die Schritte durch, die notwendig sind, um die Distanz zu berechnen.