

Übungen zur Vorlesung Sequenzanalyse

Universität Bielefeld, SS 2018

Dr. Daniel Dörr

<https://gi.cebitec.uni-bielefeld.de/teaching/2018summer/sa>

Übungsblatt 3 vom 24.04.2018

Abgabe 03.05.2018

Aufgabe 1 (Dynamische Programmierung)

(8 Punkte)

1. Schreibe zwei Haskellprogramme, die als Eingabe zwei Strings x und y bekommen und deren Edit-Distanz ausgeben. Die Edit-Distanz sollst du (a) einmal rekursiv und (b) einmal mittels dynamischer Programmierung bestimmen. Gib außerdem die Zeit aus, die jeweils für (a) und (b) benötigt wird.
2. Teste deine Programme nun mit Strings der Länge 5, 10 und 15. Trage die Zeiten, die benötigt werden, grafisch gegeneinander auf.
3. Warum ist es ineffizient, die Edit-Distanz rekursiv zu berechnen?

Aufgabe 2 (Sequenzen mit gleichem Q-Gramm Profil)

(4 Punkte)

Gegeben sei String $x = \text{TTTHTTHHTT}$ und $q = 3$.

1. Konstruiere den de Bruijn Graph $B(x, q - 1)$.
2. Gibt es eine weitere Sequenz y mit $d_q(x, y) = 0$ für $q = 3$? Wenn ja, bestimme sie, andernfalls begründe warum es eine solche Sequenz nicht geben kann.
3. Warum ist ein de Bruijn Graph immer balanciert und verbunden?

Aufgabe 3 (Die Maximal-Matches-Distanz)

(4 Punkte)

Gegeben seien die beiden Strings $x = \text{TATCTCTCT}$ und $y = \text{AATCATAT}$.

1. Berechne die Links-Rechts- und die Rechts-Links-Partition von x in Bezug zu y und gib jeweils ihren Wert an.
2. Berechne $|P_{lr}(y, x)|$ und $|P_{rl}(y, x)|$ und gib jeweils auch die Partitionen an.
3. Wie ist die Maximal-Matches-Distanz von x zu y definiert?
4. Zeige an einem von dir ausgedachten Beispiel, dass die Maximal-Matches-Distanz keine Metrik ist.