

Übungen zur Vorlesung Sequenzanalyse

Universität Bielefeld, SS 2019

Dr. Daniel Dörr · Michel T. Henrichs

<https://gi.cebitec.uni-bielefeld.de/teaching/2019summer/sa>

Übungsblatt 3 vom 22.4.2019

Abgabe am 29.4.2019 bis 12:00 Uhr (mittags)

Aufgabe 1 (Berechnung der Edit-Distanz)

(5 Punkte)

Gegeben seien die Strings $x = \text{BANANE}$ und $y = \text{TANNE}$. Benutze zur Berechnung der Edit-Distanz der beiden Strings eine Edit-Matrix und gib die Distanz an. Benutze außerdem eine weitere Matrix, in der die optimalen Edit-Operationen gespeichert werden, um alle optimalen Edit-Sequenzen bestimmen zu können. Schreibe eine optimale Edit-Sequenz explizit auf.

Aufgabe 2 (Rank und Unrank)

(3 Punkte)

Lies dir auf den Seiten 24 bis 26 im Skript durch, wie das q -Gramm Profil einer Sequenz praktisch berechnet werden kann und beantworte die folgenden Fragen.

1. Was ist eine Ranking-Funktion im Kontext der q -Gramm Distanz und was ist ihre Ein- und Ausgabe?
2. Was ist in diesem Kontext Ein- und Ausgabe einer Unranking-Funktion?
3. Wie kann man das q -Gramm Profil einer Sequenz x mit Länge n in $\mathcal{O}(n)$ statt $\mathcal{O}(q \cdot n)$ Schritten berechnen?

Aufgabe 3 (Implementierung der q -Gramm Distanz)

(7 Punkte)

Schreibe ein Programm in Java oder Python, das die q -Gramm Distanz zweier DNA-Strings berechnet. Benutze dafür das Alphabet $\Sigma = \{\text{A,C,G,T}\}$ und die aufsteigende Variante der Codierung von 0 nach $q - 1$. Gehe dabei wie folgt vor:

1. Implementiere eine Funktion `rank(x, q)`, die den Rang der ersten q Buchstaben des Wortes x berechnet.
2. Implementiere eine Funktion `rank_update(j,b,q)`, die, gegeben q , den Rang j des vorigen q -Gramms und den folgenden Buchstaben b , den Rang des neuen q -Gramms in konstanter Zeit berechnet.
3. Benutze diese beiden Funktionen, um in einer Funktion `make_profile(x, q)` das q -Gramm Profil eines Strings x in Form einer Liste oder eines Arrays zu berechnen.
4. Schreibe nun eine Funktion `q_gram_dist(x, y, q)`, die die q -Gramm Distanz der Strings x und y berechnet.
5. Gib mithilfe deines Programms die q -Gramm Distanz von $x = \text{ATTATTATTAT}$ und $y = \text{TAATAATATTAA}$ für $q = 3$ an.

Hinweis: Überlege oder suche dir einfache Beispiele zum Testen deiner Funktionen. Solltest du `rank_update` nicht implementieren können, kannst du das q -Gramm Profil auch nur über `rank` berechnen. Sollte dein Programm nicht funktionieren, berechne die Distanz aus Aufgabenteil 5 per Hand.