

Präsenzübungen zur Vorlesung Sequenzanalyse

Universität Bielefeld, WS 2013/2014

Prof. Dr. Jens Stoye · Linda Sundermann

<http://wiki.techfak.uni-bielefeld.de/gi/Teaching/2013winter/SequenzAnalyse>

Präsenzübungsblatt 4, Woche 46/2013

Aufgabe 1 (q -gram Profile und Distanzen)

Gegeben sind das Alphabet $\Sigma = \{0, 1\}$ und die Wortlänge $q = 4$.

1. Berechne die q -gram Profile der Worte $x = 010010100$ und $y = 01000101100$.
2. Was ist die Distanz von x und y .
3. Gib ein Wort $y' \neq y$ an, dass eine Distanz von 0 zu y hat.

Aufgabe 2 (Rank und Unrank)

Gegeben sind das Alphabet $\Sigma = \{A, C, G, T\}$ und die Wortlänge $q = 4$.

1. Berechne den Rang des Wortes **ATTC** und danach (ohne vollständige Neuberechnung, sondern durch ein Update in konstanter Zeit) den Rang des Wortes **TTCG**. Verwende die aufsteigende Variante der Codierung von 0 nach $q - 1$. Gib alle Zwischenschritte an.
2. Welches Wort hat den Rang 137 unter Verwendung der aufsteigenden Codierung?

Aufgabe 3 (Maximal-Matches Distanz, Eigenschaften einer Metrik)

Gegeben sind die Strings $x = \text{babbabbabaaa}$, $y = \text{bab}$ und $z = \text{ababbabbbaaaa}$.

1. Zeige für x , y und z , dass die Dreiecksungleichung für die Maximal-Matches Distanz gilt.
2. Berechne die Edit-Distanz mit Einheitskosten aller Paare aus x , y und z .
3. Welcher Zusammenhang besteht zwischen Maximal-Matches Distanz und der Edit-Distanz mit Einheitskosten?