

Übungen zur Vorlesung Sequenzanalyse I

Universität Bielefeld, Wintersemester 2010/2011
Dipl.-Inform Peter Husemann · Dr. Roland Wittler

<http://wiki.techfak.uni-bielefeld.de/gi/Teaching/2010winter/SequenzAnalyse>

Blatt 5 vom 12.11.2010

Abgabe in einer Woche vor Beginn der Vorlesung.

Aufgabe 1 Backtrac(k)ing

(2 Punkte)

Erkläre in 2–5 Sätzen den Unterschied zwischen *backtracing* und *backtracking* an dem Anwendungsfall „Aufzählen aller optimalen Edit-Sequenzen von zwei gegebenen Strings x und y “.

Aufgabe 2 q -Gram Distanz

(4 Punkte)

- Gegeben sind $x = \text{CCCTCTTTTCTTCCCTTTTT}$ und $y = \text{TCTTTCCCCCTCTTCCTTC}$ und es sei $q = 4$. Bestimme die q -Gram Profile beider Sequenzen (Schreibe die q -Gramme dabei in lexikographischer Ordnung auf) und berechne die q -Gram Distanz $d_q(x, y)$.
- Finde für $v = \text{TCCTCTTC}$ und $q = 3$ eine Sequenz w , für die gilt $d_q(v, w) = 0$ und $w \neq v$.

Aufgabe 3 Ranking und Unranking

(3 Punkte)

Zur schnellen Berechnung eines q -Gram Profiles kann jedes q -Gram mittels einer Rankingfunktion auf eine natürliche Zahl abgebildet werden. Ein Beispiel für eine solche Funktion ist

$$r(x) = \sum_{i=1}^q r_{\Sigma}(x[i]) \cdot |\Sigma|^{i-1}$$

welche einer Sequenz $x \in \Sigma^q$ ihren Rang $r(x) \in \mathbb{N}_0^+$ zuweist. Dabei wird die Funktion $r_{\Sigma}(\cdot)$ benutzt, die das Alphabet Σ auf die Zahlen $\{0, \dots, |\Sigma| - 1\}$ abbildet.

Gegeben sei nun das Alphabet $\Sigma = \{\mathbf{A}, \mathbf{C}, \mathbf{G}, \mathbf{T}\}$ ($r_{\Sigma}(\mathbf{A}) = 0$, $r_{\Sigma}(\mathbf{C}) = 1$, ...) und die Wortlänge $q = 4$.

- Berechne den Rang des Wortes $x = \mathbf{CGAT}$. (Gib dabei den Rechenweg an.)
- Berechne, ohne vollständige Neuberechnung, sondern durch ein Update in konstanter Zeit, ausgehend von dem Rang von x den Rang des Wortes $y = x[2], x[3], x[4], \mathbf{G} = \mathbf{GATG}$. (Gib ebenfalls den Rechenweg an.)
- Welche Sequenz $z \in \Sigma^q$ hat den Rang 112? (Gib den Rechenweg an.)