
Algorithmische Bioinformatik I

Aufgabe 1

Seien $s, t \in \Sigma^*$ und $|s| = |t| = n$. Zeigen Sie, dass die Zahl der gültigen unterschiedlichen Alignments von s und t exponentiell in der Länge der Sequenzen ist. Gehen Sie dabei wie folgt vor.

- Zeigen Sie zunächst, dass $\binom{2n}{n}$ eine untere Schranke für die Anzahl der Alignments ist.
- Verwenden Sie die Formel von Stirling ($n! \approx \sqrt{2\pi n} \cdot n^n \cdot e^{-n}$ für $n \rightarrow \infty$) um zu zeigen, dass dies ein in n exponentielles Wachstum impliziert.

Aufgabe 2

Sei $s = s_1 \cdots s_n \in \Sigma^*$ eine Zeichenfolge. Eine Zeichenfolge $s' = s'_1 \cdots s'_\ell \in \Sigma^*$ heißt *Teilfolge* von s , wenn es eine streng aufsteigende Folge $(k_1, \dots, k_\ell) \in \mathbb{N}^\ell$ gibt (d.h., es gilt $k_1 < k_2 < \dots < k_\ell$), so dass $s'_i = s_{k_i}$ gilt.

Beispiel: *eure* ist eine Teilfolge von *Sequenzenvergleich*.

Seien $s = s_1 \cdots s_n \in \Sigma^*$ und $t = t_1 \cdots t_m \in \Sigma^*$ zwei Zeichenfolgen. Konstruiere einen Algorithmus, der eine längste Zeichenfolge $u = u_1 \cdots u_\ell$ findet, die sowohl eine Teilfolge von s als auch von t ist.

Aufgabe 3

Geben Sie eine auf lokalem Sequenzen-Alignment basierende Methode an, um eine optimale Wiederholung, d.h. ein Paar $(s_i \cdots s_{i'}, s_j \cdots s_{j'})$ von Teilstrings mit maximalem Score, innerhalb eines gegebenen Strings $s \in \Sigma^*$ zu finden. Das gefundene Alignment darf in der Tabelle der Dynamischen Programmierung die Diagonale nicht berühren, damit kein Buchstabe mit sich selbst aliniert wird.