

4. Übung zur Vorlesung “Einführung in die Bioinformatik I, 1. Teil”

Wintersemester 2020/2021

Prof. Peter Dittrich, Prof. Sebastian Böcker, Emanuel Barth, Marcus Ludwig

Ausgabe: 25. November 2020,
Abgabe: 02. Dezember 2020 zu Beginn der Übung

Landau-Symbole

Aufgabe 1 (5 Punkte): Vergleichen Sie jeweils mit Hilfe von Landau-Symbolen: $\log \log n$ und $\log^2 n$, $\log n$ und \sqrt{n} , $\sum_{i=1}^n i$ und $(14 \log n + \frac{n}{4})^2$, $2 + \sin n$ und 1 , $\frac{1}{n}$ und $\frac{1}{\log n}$. Wählen Sie immer das aussagekräftigste Symbol!¹

Z-Algorithmus

Aufgabe 2 (8 Punkte): Berechnen Sie für die Strings `ababcaab`, `abbababbac` und `aabcadaaab` jeweils die Werte Z_i , l_i und r_i . Beschreiben Sie dabei die Arbeitsweise des Z-Algorithmus für die drei Strings:

- Beschreiben Sie, was der Algorithmus in jedem Schleifendurchlauf (für jeden Wert von k) tut.
- Zählen Sie für jeden String die Anzahl von Matches und Mismatches (durch direkte Buchstabenvergleiche!).

Aufgabe 3 (2 Punkte): Wenn $Z_2 = q > 0$, welche Werte haben dann Z_3, Z_4, \dots, Z_{q+2} ?

Aufgabe 4 (5 Punkte): In Fall 2 des Z-Algorithmus gibt es eigentlich drei Unterfälle: $Z_{k'} < |\beta|$, $Z_{k'} = |\beta|$ und $Z_{k'} > |\beta|$. Begründen Sie, warum aus $Z_{k'} > |\beta|$ folgt, dass $Z_k = |\beta|$.

Bonusaufgabe (3 Punkte): Betrachten Sie folgendes Problem:

EINGABE: Zwei Strings s und t der Länge n , eine natürliche Zahl $k > 0$

AUSGABE: Die Anzahl identischer Positionen² in jedem Paar (s', t') , wobei s' ein Substring der Länge k von s und t' ein Substring der Länge k von t ist.

¹ $\log^2 n = (\log n)^2$

²Beispiel: In `abcde` und `abdba` sind die erste, die zweite und die vierte, insgesamt also drei Positionen identisch.

Der naive Algorithmus für dieses Problem hat eine Laufzeit von $O(k \cdot n^2)$. Finden Sie einen Algorithmus mit einer Laufzeit von $O(n^2)$, und begründen Sie dessen Korrektheit und Laufzeit.