

Algorithmen auf Sequenzen

Übung – Blatt 8

Ausgabe: 26. Mai, **Besprechung:** 31.05. 9:00 Uhr

Aufgabe 8.1

Gib alle in der Vorlesung vorgestellten Abstandsmaße für Strings und deren Definitionen an.

Aufgabe 8.2

Berechne für folgende Paare von Strings die *Edit-Distanz*:

- *banane, ananas*
- *vogel, igel*
- *jaegerboot, jamesbond*
- *rahmann, kopczynski*

Aufgabe 8.3

Definiere $d_{lcs} := \max\{|s|, |t|\} - lcs(s, t)$ und d_{lcf} entsprechend über *lcf*. Sind diese definierten Abstände Metriken?

Aufgabe 8.4

Suche im Text GCTATACGAACTTGA alle Vorkommen des Muster TACTG mit maximal 2 Fehlern mit dem *Shift-And Algorithmus* heraus.

Aufgabe 8.5

Da sich ein BWT-transformierter Text häufig besser komprimieren lässt, liegt es nahe, die Transformation mehrfach anzuwenden, also zu einem String s die Kompressionseigenschaften von s , $bwt(s)$, $bwt(bwt(s))$, usw. zu betrachten. Warum “bringt das nichts”?

Präsentationsaufgabe 8.6 :

Implementiere die BWT und Backward Search (zumindest mit fester Alphabetgröße 4) und suche damit DNA-Sequenzen verschiedener Länge im Text *chromosom.txt*¹. Wie verhalten sich Suchzeit und Speicherbedarf für verschiedene Samplingraten k für Occ?

¹<http://ls11-www.cs.tu-dortmund.de/people/kopczyns/Lehre/AAS/chromosom.txt>