

Übungen zur Vorlesung Sequenzanalyse II

Universität Bielefeld, SS 2011
Prof. Dr. Jens Stoye · Daniel Dörr

<http://wiki.techfak.uni-bielefeld.de/gi/Teaching/2011summer/SequenzAnalyse>

Blatt 3 vom 13.05.2011

Bitte gib auch den Namen deiner Tutorin an.

Abgabe am 24.05.2011 zu Beginn der Vorlesung

Aufgabe 1 (Suffixarray)

4 Punkte

Gegeben ist der String $s\$ = \text{GCATAAATAAA}\$, wobei $\$ < A < C < G < T$.$

- Berechne mit Hilfe des Manber-Myers-Algorithmus das Suffixarray pos für $s\$$. Gib dabei für jede Phase das Array und die Grenzen der *buckets* an (Kennzeichne wie folgt: $||$ für Grenzen der vorangegangenen Phasen, $|$ für neue Grenzen der jeweiligen aktuellen Phase). Wende den Algorithmus solange an, bis alle Suffixe in pos sortiert sind.
- Gib das **rank**- und das **lcp**-Array von $s\$$ an.

Aufgabe 2 (Paarweises Alignment mit linearem Speicherbedarf)

4 Punkte

Geben seien zwei Strings $s = \text{TGGATCAGAG}$ und $t = \text{AGGCTCGAGA}$. Berechne das globale Alignment von s und t unter Verwendung der Divide-and-Conquer Technik. Simuliere die dazu notwendigen Schritte auf Kopien der Edit-Matrix:

- Kennzeichne jeweils $m' = \lceil m/2 \rceil$ und zeichne in der unteren Hälfte der Edit-Matrix die Backtracing-Pfeile ein.
- Gib jeweils die aktuelle Zeile der Matrix M an. Diese enthält die Zellen der mittleren Zeile m' , durch die ein optimales Alignment verlaufen würde.
- Zeige in jedem Rekursionsschritt, welche Teile der Edit-Matrix erneut berechnet werden müssen.

Aufgabe 3 (Divide-and-Conquer (D&C) Technik)

4 Punkte

Mittels der D&C Technik lassen sich paarweise Alignments mit linearem Speicherbedarf berechnen. In der Vorlesung wurde ein Algorithmus vorgestellt, mit welchem das optimale globale Alignment bestimmt werden kann. Welche Probleme treten auf, wenn die D&C Technik zur Berechnung folgenden Alignment-Varianten angewendet wird, und wie lassen sie sich lösen?

- globales Alignment mit affinen Gapkosten (Gotoh)*
Bei affinen Gapkosten wird zwischen Kosten für zum Öffnen (*gap-open*) und Erweitern (*gap-extension*) von Gaps unterschieden.
- lokales Alignment (Smith-Waterman)*
Lokales Alignment ist das beste globale Alignment aller Substrings zweier Sequenzen.
- Ausgabe von k (sub-) optimalen Alignments (Waterman-Eggert)*
In einem Alignment wird Backtracking zum Auffinden eines optimalen Alignments angewandt. Häufig gibt es jedoch mehrere optimale Alignments, und in manchen Fällen ist es auch interessant, die k ($k = 1, 2, 3, \dots$) besten – möglicherweise suboptimalen – nicht-überlappenden Alignments zu kennen.