

Übungen zur Vorlesung Sequenzanalyse II

Universität Bielefeld, SS 2010
Prof. Dr. Jens Stoye · Dr. Inke Herms

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

Blatt 3 vom 30.04.2010

**Abgabe in einer Woche vor Beginn der Vorlesung.
Bitte gib auch den Namen deines Tutors an.**

Aufgabe 1 Suffixarray

3 Punkte

Gegeben ist der String $s = \text{GTAACAGTACA}$. Gib das Suffixarray `pos`, sein Inverses `rank` und das lcp-Array von `s` an. Beachte $\$ < A < C < G < T$ und Indizierung beginnend mit 0.

Aufgabe 2 Suffixarray Interpretation

3 Punkte

Inwiefern ist das Suffixarray `pos` ein q -gram Index für alle $q \geq 1$ gleichzeitig?

Aufgabe 3 Längster gemeinsamer Substring

3 Punkte

Gib einen Linearzeit-Algorithmus an, mit dem man einen längsten gemeinsamen Substring zweier Strings s und t finden kann, wenn das Suffixarray von $s\#t$ gegeben ist.

Aufgabe 4 Manber-Myers-Algorithmus

5 Punkte

Implementiere den Manber-Myers-Algorithmus in seiner einfachsten Form:

```
int[] posManberMyers (byte[] s, int alphabetsize) {...}
```

Dabei stellt `s` den Eingabetext in Form eines byte-Arrays dar und `alphabetsize` die Alphabetgröße. Es kann davon ausgegangen werden, dass `s[0..|s|-2]` nur aus Bytes zwischen 0 und `alphabetsize-1` besteht und dass `s[|s|-1] == alphabetsize`. Ausgabe ist das Suffixarray `pos`.