

Übungen zur Vorlesung Sequenzanalyse II

Universität Bielefeld, SoSe 2007

Dr. Sven Rahmann · Dipl.-Inf. Peter Husemann · Dipl.-Biol. Constantin Bannert

<http://gi.cebitec.uni-bielefeld.de/teaching/2007summer/sequenzanalyse/>

Blatt 1 vom 20.04.2007

Abgabe am 27.04.2007 vor der Vorlesung um 8:30 in H14

Aufgabe 1 In den inneren Knoten eines Suffixbaumes müssen auf irgendeine Weise die Kinder gespeichert werden. Wie viel Platz insgesamt (Textlänge n und Alphabetgröße σ) und wie viel Zeit zum Auffinden einer bestimmten ausgehenden Kante benötigt man jeweils bei Verwendung

1. eines Arrays fester Größe σ ?
2. eines sortierten Arrays aller Kinder?
3. einer verketteten Liste aller Kinder?
4. eines balancierten Suchbaumes aller Kinder?
5. einer Hash-Tabelle aller Kinder?

Aufgabe 2 Sei v innerer Knoten mit $\text{string}(v) = ax$ mit $a \in \Sigma$ und $x \in \Sigma^*$; d.h. $\text{node}(ax)$ existiert und ax ist rechtsverzweigend. Zeige, dass dann auch $\text{node}(x)$ existiert.

Bemerkung: Dies erlaubt es, stets sogenannte Suffixlinks zu definieren, die von einem Knoten $\text{node}(ax)$ auf den Suffixknoten $\text{node}(x)$ zeigen. Es ist ja nicht selbstverständlich, dass so ein Knoten existiert!

Aufgabe 3 Veranschauliche das Vorgehen des WOTD-Algorithmus für den String `bacbacba$`.

Aufgabe 4 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.length-2]` nur aus Bytes zwischen 0 und `alphabetsize-1` besteht und dass `s[s.length-1] == alphabetsize`. Ausgabe ist das Suffix-Array `pos`.