

Algorithmen zur Sequenzanalyse: Übungsblatt 4

Enno Ohlebusch
Timo Beller

Das Übungsblatt wird am 10.12.2013 besprochen.

Aufgabe 1

Ein String ω ist ein *prefix tandem repeat* des Strings S wenn gilt:

- ω ist ein Präfix von S
- es gibt einen String u mit $uu = \omega$

Geben Sie einen linearen Algorithmus an, welcher das längste *prefix tandem repeat* von S findet.

Aufgabe 2

Sei ω ein längstes Repeat. Zeigen Sie, dass ω ein supermaximales Repeat (und damit auch ein maximales Repeat) sein muss.

Aufgabe 3

Zeichnen Sie einen de Bruijn Graph für $k = 2$ über dem Alphabet $\Sigma = \{a, c, g, t\}$. Finden Sie einen Eulerkreis in dem Graph und geben Sie die korrespondierende nicht-zyklische de Bruijn Sequenz an. Wie viele supermaximale Repeats enthält die Sequenz?

Aufgabe 4

Zeigen Sie, dass Algorithmus 1 eine lineare Laufzeit hat.

Aufgabe 5

Entwerfen Sie einen linearen Algorithmus welcher den *lcp*-Intervalbaum Bottom-Up durchläuft und dabei simultan alle maximalen und supermaximalen Repeats ausgibt.

Algorithm 1 Berechnung supermaximaler Repeats.

```
 $\ell \leftarrow -1$   
 $locMax \leftarrow false$   
for  $k \leftarrow 2$  to  $n + 1$  do  
   $m \leftarrow LCP[k]$   
  if  $m > \ell$  then  
     $lb \leftarrow k - 1$   
     $locMax \leftarrow true$   
  else if  $m < \ell$  and  $locMax = true$  then  
     $rb \leftarrow k - 1$   
     $pd \leftarrow true$   
    for  $i \leftarrow lb$  to  $rb$  do  
       $c \leftarrow BWT[i]$   
      if  $B[c] = 1$  then  
         $pd \leftarrow false$   
         $B[c] \leftarrow 1$   
      if  $pd = true$  then  
        report that  $\ell - [lb..rb]$  induces a supermaximal repeat  
    for  $i \leftarrow lb$  to  $rb$  do  
       $c \leftarrow BWT[i]$   
       $B[c] \leftarrow 0$   
     $locMax \leftarrow false$   
 $\ell \leftarrow m$ 
```
