Algorithmen und Datenstrukturen II

SoSe 2010, **4. Aufgabenblatt**, Abgabe 09.06.2010

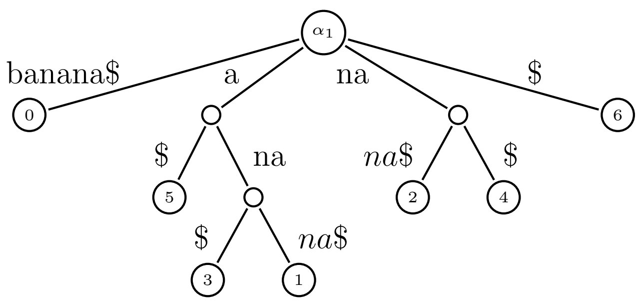
Aufgabe 10

12 Punkte

Bauen Sie einen Suffixbaum für den String `banana$` nach der vorgestellten naiven Konstruktionsmethode. Verwenden Sie die Notation aus der Vorlesung. Geben Sie analog zum dortigen Beispiel die Abfolge aller Funktionsaufrufe (`addLeaf`, `insertEdge`, `getlcp`) innerhalb von `topdown` an. Verdeutlichen Sie die Rekursionsstufen durch Einrückungen. Zeichnen Sie den Baum.

$U := \{(banana\$, 0), (anana\$, 1), (nana\$, 2), (ana\$, 3), (na\$, 4), (a\$, 5), (\$, 6)\}$

- $P_a = \{(nana\$, 1), (na, 3), (\$, 5)\}, getlcp(P_a) = 0$
  - $P_n = \{(ana\$, 1), (a\$, 3)\}, getlcp(P_n) = 1$
    - $P_n = \{a\$, 1\} \longrightarrow addLeaf(\alpha_3, na\$, 1)$
    - $P_\$ = \{\epsilon, 3)\} \longrightarrow addLeaf(\alpha_3, \$, 3)$
  - $\longrightarrow insertEdge(\alpha_2, na, \alpha_3)$
  - $P_\$ = \{(\$, 5)\} \longrightarrow addLeaf(\alpha_2, \$, 5)$
- $\longrightarrow insertEdge(\alpha_1, a, \alpha_2)$
- $P_b = \{(anana\$, 0)\}, \longrightarrow addLeaf(\alpha_1, \$, 0)$
- $P_n = \{(ana\$, 2), (a\$, 4)\}, getlcp(P_n) = 1$
  - $P_n = \{(a\$, 2)\}, \longrightarrow addLeaf(\alpha_4, na\$, 2)$
  - $P_\$ = \{\epsilon, 4\} \longrightarrow addLeaf(\alpha_4, \$, 4)$
- $\longrightarrow insertEdge(\alpha_1, na, \alpha_4)$
- $P_\$ = \{(\epsilon, 6)\}, \longrightarrow addLeaf(\alpha_1, \$, 6)$



Aufgabe 11

6 Punkte

Füer einen beliebigen String s endlicher Länge bezeichnen wir mit $\mathrm{Sub}(s)$ die Menge aller Substrings von s.

(a) Zeigen Sie, dass es zu jeder natürlichen Zahl n ein Alphabet $\mathcal{A}$ und einen String s der Länge n über dem Alphabet $\mathcal{A}$ gibt, so dass

$$|\,\mathrm{Sub}(s)| = n(n+1)/2\ .$$

(3 Punkte)

Wähle das Alphabet $\mathcal{A} = \{1, 2, \ldots, n\}$ und $s[\,] = 1\,2\,3\,\ldots\,n$. Betrachte die Menge von Indexpaaren $I := \{(i, j) : 1 \leq i \leq j \leq n\}$. Für $(i, j) \in I$ und $(k, l) \in I$ impliziert $(i, j) \neq (k, l)$, dass $s[i \ldots j] \neq s[k \ldots l]$. (Die beiden Substrings unterscheiden sich im ersten oder im letzten Zeichen). Somit haben wir eine injektive Abbildung von I auf $\mathrm{Sub}(s)$ gefunden. Andererseits ist diese Abbildung nach Konstruktion surjektiv, es werden also alle Substrings getroffen. Wir finden also

$$|\,\mathrm{Sub}(s)| = |I| = \sum_{i=1}^{n} i = \frac{n(n+1)}{2}\ .$$

(b) Seien $\mathcal{A}$ und $\mathcal{B}$ Alphabete mit $\mathcal{A} \cap \mathcal{B} = \emptyset$. Seien s ein String über $\mathcal{A}$ und t ein String über $\mathcal{B}$. Betrachten Sie nun die Konkatenation st, die ein String über dem Alphabet $\mathcal{A} \cup \mathcal{B}$ ist. Wie können Sie $|\,\mathrm{Sub}(st)|$ aus $|s|$, $|t|$, $|\,\mathrm{Sub}(s)|$ und $|\,\mathrm{Sub}(t)|$ berechnen? Begründen Sie Ihre Antwort. (3 Punkte)

$$|\,\mathrm{Sub}(st)| = |\,\mathrm{Sub}(s)| + |\,\mathrm{Sub}(t)| + |s| \times |t|$$

Der erste Summand zählt die Substrings von st, die ganz in s liegen. Der zweite Summand diejenigen, die ganz in t liegen. Da die Alphabete disjunkt sind, sind diese Mengen ebenfalls disjunkt. Die restlichen Substrings von st beginnen in s und enden in t, sind also durch die Wahl einer von $|s|$ Anfangspositionen und einer von $|t|$ Endpositionen eineindeutig bestimmt.

Aufgabe 12

10 Punkte

Gegeben sind der Text

$$x[\,] = \texttt{AMMONIUMNITRATKRATERKAMMER}$$

und die Muster

$$y_1[\,] = \texttt{MAMA},\ y_2[\,] = \texttt{AMME}\ .$$

(a) Wenden Sie den Algorithmus von Knuth-Morris-Pratt an, um in x nach y_1 zu suchen. Geben Sie dazu zuerst die *next*-Tabelle für y_1 an. Beschreiben Sie den Verlauf der Suche, indem Sie die Paare (i, j) in einem Diagramm auftragen wie in der Vorlesung vorgeführt. Beschriften Sie die Achsen und zeichnen Sie so präzise, daß die angenommenen (i, j)-Paare eindeutig ablesbar sind. Wieviele Vergleiche zwischen Text und Muster werden ausgeführt? (4 Punkte)

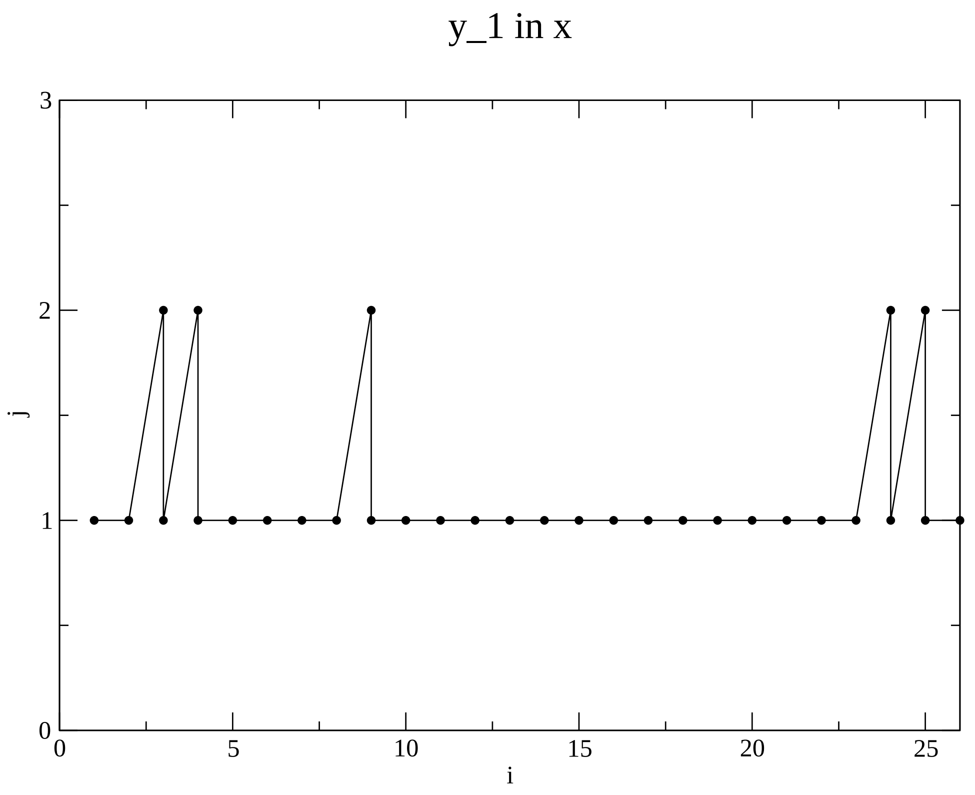
(b) Verwenden Sie den Algorithmus von Boyer-Moore (laut Vorlesung), um in t nach y_1 und nach y_2 zu suchen. Geben Sie für jedes dieser beiden Muster die *last*-Tabelle an. Notieren Sie die Folge der Textpositionen i, an die das Muster “angelegt” wird. (6 Punkte)

```
(a) Knuth-Morris-Pratt

j       1 2 3 4
g[j]    M A M A
next[j] 0 1 1 2

1  2  3  4  5  6  7  8  9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26
A  M  M  O  N  I  U  M  N  I  T  R  A  T  K  R  A  T  E  R  K  A  M  M  E  R
M  A  M  A                             M  A  M  A
   M  A  M  A                             M  A  M  A
      M  A  M  A                             M  A  M  A
         M  A  M  A                             M  A  M  A
            M  A  M  A                             M  A  M  A
               M  A  M  A                             M  A  M  A
                  M  A  M  A                             M  A  M  A
                     M  A  M  A                             M  A  M  A
                        M  A  M  A                             M  A  M  A
                           M  A  M  A                             M  A  M  A
                              M  A  M  A                             M  A  M
                                 M  A  M  A                             M  A
                                    M  A  M  A                             M

31 Vergleiche
```



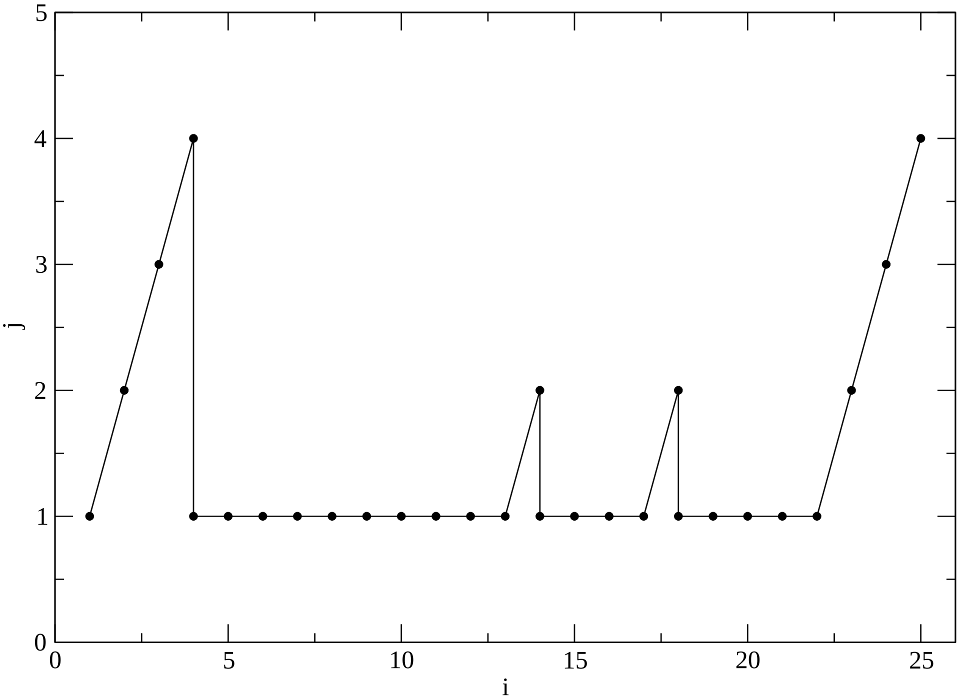
```
Zusatz:

j       1 2 3 4
g[j]    A M M E
next[j] 0 1 1 1

1  2  3  4  5  6  7  8  9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26
A  M  M  O  N  I  U  M  N  I  T  R  A  T  K  R  A  T  E  R  K  A  M  M  E  R
A  M  M  E                          A  M  M  E
         A  M  M  E                    A  M  M  E
            A  M  M  E                    A  M  M  E
               A  M  M  E                    A  M  M  E
                  A  M  M  E                    A  M  M  E
                     A  M  M  E                    A  M  M  E
                        A  M  M  E                    A  M  M  E
                           A  M  M  E                    A  M  M  E
                              A  M  M  E                    A  M  M  E
                                 A  M  M  E                    A  M  M  E

28 Vergleiche
```

y_2 in x



(b) Boyer-Moore

y_1 in x:

Last-Tabelle:

0	1	2	3
M	A	M	A

also

char	last appearance
A	3
M	2
sonst	-1

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26
A	M	M	O	N	I	U	M	N	I	T	R	A	T	K	R	A	T	E	R	K	A	M	M	E	R
M	A	M	A																						
				M	A	M	A																		
					M	A	M	A																	
									M	A	M	A													
												M	A	M	A										
																M	A	M	A						
																				M	A	M	A		
																					M	A	M	A	

also $i = 1, 5, 6, 10, 13, 17, 21, 22$

y_2 in x:

Last-Tabelle:

0	1	2	3
A	M	M	E

also

char	last appearance
A	0
E	3
M	2
sonst	-1

1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	26
A	M	M	O	N	I	U	M	N	I	T	R	A	T	K	R	A	T	E	R	K	A	M	M	E	R
A	M	M	E																						
				A	M	M	E																		
					A	M	M	E																	
									A	M	M	E													
												A	M	M	E										
																A	M	M	E						
																				A	M	M	E		
																					A	M	M	E	

also $i = 1, 5, 6, 10, 13, 17, 21, 22$