

Stringmatching

Prüfungsvorbereitung

Aufgabe 1

Führe eine Suche nach dem Muster ADA im Text YABBADABBAD00 mit der naiven (brute force) Methode. durch. Beschreibe die Schritte detailliert und ermittle die Gesamtzahl der Vergleiche für beide Algorithmen.

Aufgabe 1

Naive Methode:

Y	A	B	B	A	D	A	B	B	A	D	O	O	Vergleiche
A	D	A											1
	A	D	A										2
		A	D	A									1
			A	D	A								1
				A	D	A							3
													8

Aufgabe 2

Wie viele Vergleiche sind nötig, um mit dem naiven Verfahren herauszufinden, dass der Text AAAA...A aus 1000 Symbolen A das Muster AAAAB nicht enthält?

Aufgabe 2

$$(1000 - 5 + 1) \cdot 5 = 5 \cdot 996 = 4980 \text{ Vergleiche}$$

Aufgabe 3

Wie viele Vergleiche sind nötig, um mit dem Verfahren von Boyer-Moore-Horspool herauszufinden, dass der Text AAAA...A aus 1000 Symbolen A das Muster AAAAB nicht enthält?

Aufgabe 3

Verfahren von Boyer-Moore-Horspool:

Bad Character Table:

A	B
1	5

$(1000 - 5 + 1) \cdot 1 = 996$ Vergleiche

Aufgabe 4

- (a) Erzeuge den deterministischen endlichen Automaten (DFA), mit dem nach dem Muster GGCA gesucht werden kann.
- (b) Zeige, schrittweise, wie die Suche im Text GGGAAAGGCAT mit dem in (a) konstruierten DFA funktioniert und wie die Position eines Treffers bestimmt wird.

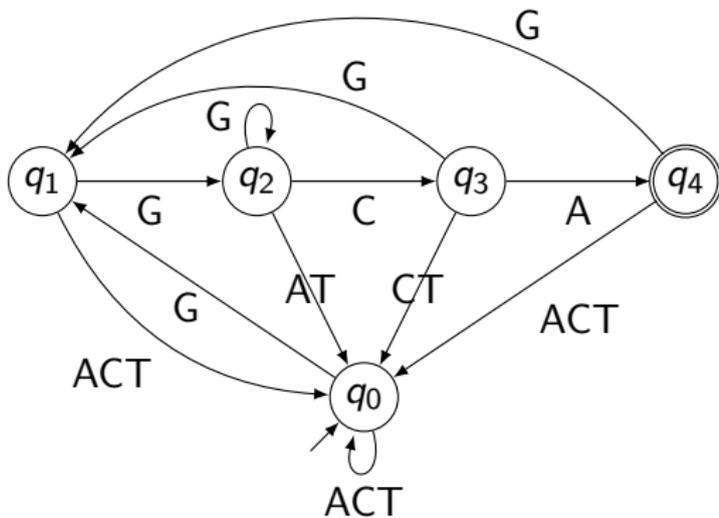
Aufgabe 4

(a) s : längstes Suffix von $p[0:i]+x$, das Präfix von $p = \text{GGCA}$ ist.

q_i	$p[0:i]+x$	s	q_{i+1}
0	A	ε	0
0	C	ε	0
0	G	G	1
0	T	ε	0
1	GA	ε	0
1	GC	ε	0
1	GG	GG	2
1	GT	ε	0
2	GGA	ε	0
2	GGC	GGC	3
2	GGG	GG	2
2	GGT	ε	0

q_i	$p[0:i]+x$	s	q_{i+1}
3	GGCA	GGCA	4
3	GGCC	ε	0
3	GGCG	G	1
3	GGCT	ε	0
4	GGCAA	ε	0
4	GGCAC	ε	0
4	GGCAG	G	1
4	GGCAT	ε	0

Der DFA in graphischer Darstellung:



(b)

Index:	-	0	1	2	3	4	5	6	7	8	9	0	1
Zeichen:	ϵ	G	G	G	A	A	A	G	G	C	A	T	
Zustand:	0	1	2	2	0	0	0	1	2	3	4	0	
Match:	-	-	-	-	-	-	-	-	-	-	6	-	

Position des Treffers: $j = 9 - \text{len}(p) + 1 = 9 - 4 + 16$

Aufgabe 5

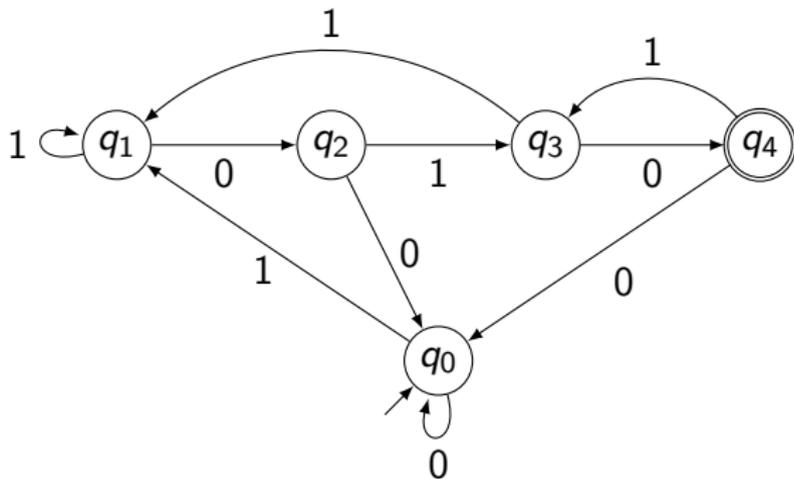
- (a) Erzeuge den deterministischen endlichen Automaten (DFA), mit dem in einem Text über dem Alphabet $\Sigma = \{0, 1\}$ nach dem Muster 1010 gesucht werden kann.
- (b) Zeige, schrittweise, wie die Suche im Text 10110101010 mit dem in (a) konstruierten DFA funktioniert und wie die Position eines Treffers bestimmt wird.

Aufgabe 5

- (a) Bestimme für alle $0 \leq i < \text{len}(p)$ und alle $x \in \Sigma$ jeweils das längste Suffix von $p[0:i] + x$, das Präfix von $p = 1010$ ist.

q_i	$p[0:i]+x$	s	q_{i+1}
q_0	0	ε	q_0
q_0	1	1	q_1
q_1	10	10	q_2
q_1	11	1	q_1
q_2	100	ε	q_0
q_2	101	101	q_3
q_3	1010	1010	q_4
q_3	1011	1	q_1
q_4	10100	ε	q_0
q_4	10101	101	q_3

Der DFA in graphischer Darstellung:



(b)

Index:	-	0	1	2	3	4	5	6	7	8	9	0	1
Zeichen:	ε	1	0	1	1	0	1	0	1	0	1	0	
Zustand:													
Match:	-	-	-	-	-	-	-	-	-	-	6	-	

Position des Treffers: $j = 9 - \text{len}(p) + 1 = 9 - 4 + 16$