

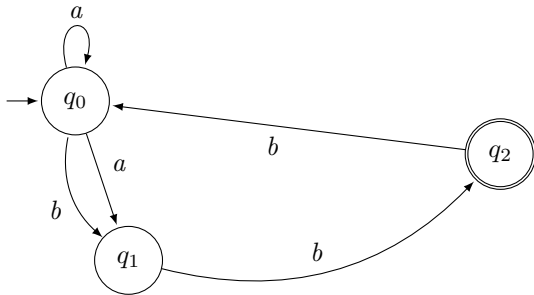
Einführung in die Computerlinguistik Zwischenklausur

Laura Kallmeyer

Heinrich-Heine-Universität Düsseldorf

Erlaubte Hilfsmittel: Eine Din-A4 Seite mit Notizen.

Aufgabe 1 Betrachten Sie den folgenden NFA:



1. Geben Sie den Automaten in Tupelschreibweise an. D.h., als Tupel $\langle Q, \Sigma, \delta, q_0, F \rangle$.
2. In welchen Zustandsmengen befindet sich der Automat nach den folgenden Übergängen?
 - (a) $\hat{\delta}(q_0, a)$ (b) $\hat{\delta}(q_2, ab)$ (c) $\hat{\delta}(q_1, bbaab)$

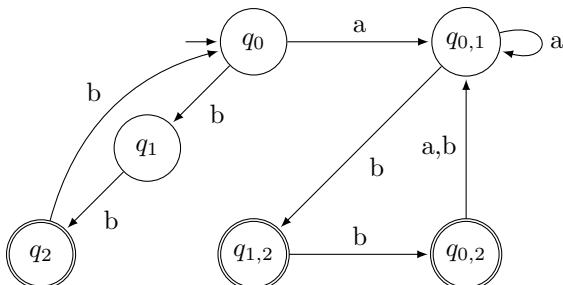
Lösung:

1. $\langle \{q_0, q_1, q_2\}, \{a, b\}, \delta, q_0, \{q_2\} \rangle$ mit

$\delta(q_0, a) = \{q_0, q_1\}$	$\delta(q_0, b) = \{q_1\}$	3 Pkte
$\delta(q_1, a) = \emptyset$	$\delta(q_1, b) = \{q_2\}$	
$\delta(q_2, a) = \emptyset$	$\delta(q_2, b) = \{q_0\}$	
2. (a) $\hat{\delta}(q_0, a) = \{q_0, q_1\}$ (b) $\hat{\delta}(q_2, ab) = \emptyset$ (c) $\hat{\delta}(q_1, bbaab) = \{q_1, q_2\}$ 3 Pkte

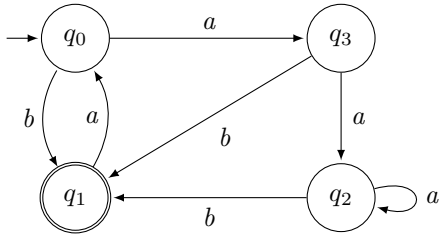
Aufgabe 2 Betrachten Sie nochmals den NFA aus der vorhergehenden Aufgabe. Geben Sie einen äquivalenten DFA an.

Lösung:



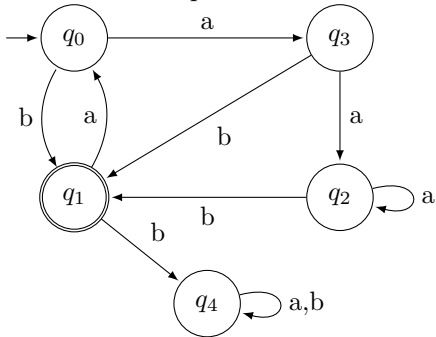
7 Pkte

Aufgabe 3 Minimieren Sie den folgenden DFA. Geben Sie dabei die Ergebnisse der einzelnen Schritte (Automat mit trap state, Matrix, minimierter Automat) an.



Lösung:

Automat mit Trap State:



2 Pkte

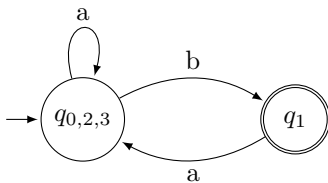
$|Q| \times |Q|$ -Matrix:

	0	1	2	3
4	X	X	X	X
3		X		
2		X		
1	X			

4 Pkte

Entferne anschließend wieder den überflüssige Trap State.

Minimierter Automat (q_0, q_2 und q_3 sind äquivalent):



3 Pkte

Aufgabe 4

1. Welche Sprachen werden von den folgenden regulären Ausdrücken denotiert?

(a) $a(b|c)$ (b) $ab|c^d^*$ (c) $(a(a|b))^+$

2. Geben Sie für folgende Sprachen jeweils einen entsprechenden regulären Ausdruck an:

(a) $\{a, b, ab, c\}$

(b) $\{w \in \{a, b\}^* \mid \text{jedes } a \text{ in } w \text{ ist notwendig von einem } b \text{ gefolgt}\}$

(c) $\{w \in \{a, b\}^* \mid w \text{ enthält entweder nur } as, \text{ und zwar eine gerade Anzahl, oder } w \text{ enthält nur } bs, \text{ und zwar eine ungerade Anzahl}\}$

Lösung:

1. (a) $L(a(b|c)) = \{ab, ac\}$

1 Pkt

- (b) $L(ab|c\epsilon d^*) = \{ab\} \cup \{cw \mid w \text{ ist eine beliebig lange, eventuell auch leere Folge von } ds\}$ 2 Pkte
 (c) $L((a|b)^+) = \{w_1 \dots w_n \mid n \geq 1, w_i \in \{aa, ab\} \text{ f\"ur } 1 \leq i \leq n\}$ 2 Pkte
2. (a) $(a|b|ab|c)$ 1 Pkt
 (b) $(b|ab)^*$ 2 Pkte
 (c) $(aa)^*|b(bb)^*$ 3 Pkte

Aufgabe 5

1. Welche Sprache wird von der folgenden linkslinearen Grammatik erzeugt?

$G = \langle \{S, A, B\}, \{a, b, c\}, P, S \rangle$ mit folgenden Produktionen in P :

$S \rightarrow Aa, S \rightarrow B, A \rightarrow Aa, A \rightarrow a, B \rightarrow Bbc, B \rightarrow \epsilon$

2. Geben Sie eine rechtslineare Grammatik für folgende Sprache an: $\{w \in \{a, b\}^* \mid \text{jedes } a \text{ in } w \text{ ist notwendig von einem } b \text{ gefolgt}\}$

Lösung:

1. $\{a^n \mid n \geq 2\} \cup \{(bc)^m \mid m \geq 0\}$ 4 Pkte
2. Grammatik: $\langle \{S, A\}, \{a, b\}, P, S \rangle$ mit folgenden Produktionen in P :
 $S \rightarrow bS, S \rightarrow aA, A \rightarrow bS, A \rightarrow \epsilon$ 3 Pkte

Aufgabe 6 Nehmen Sie an, Sie haben einen HMM-POS Tagger, unter anderem mit folgenden Wahrscheinlichkeiten (Pro, V, N sind mögliche POS-Tags):

Emissionswahrscheinlichkeiten:

$$P(\text{they}|\text{Pro}) = 5 \cdot 10^{-3} \quad P(\text{promise}|\text{N}) = 2 \cdot 10^{-3} \quad P(\text{help}|\text{N}) = 3 \cdot 10^{-3}$$

$$P(\text{promise}|\text{V}) = 4 \cdot 10^{-3} \quad P(\text{help}|\text{V}) = 2 \cdot 10^{-3}$$

Alle anderen Emissionswahrscheinlichkeiten für they, promise und help seien 0.

Übergangswahrscheinlichkeiten sind (unter anderem):

$$P(V|\text{Pro}) = 6 \cdot 10^{-1} \quad P(N|\text{Pro}) = 1 \cdot 10^{-1}$$

$$P(V|V) = 2 \cdot 10^{-1} \quad P(N|V) = 3 \cdot 10^{-1}$$

$$P(V|N) = 4 \cdot 10^{-1} \quad P(N|N) = 2 \cdot 10^{-1}$$

Die Wahrscheinlichkeiten, dass der Satz mit einem Pro eingeleitet wird, ist $1 \cdot 10^{-1}$. Die Wahrscheinlichkeit, dass auf ein N oder V ein Satzende folgt, ist jeweils $0.1 = 1 \cdot 10^{-1}$.

1. Geben Sie die Viterbi Matrix an, die sich bei diesen Wahrscheinlichkeiten für die Eingabe they promise help ergibt. Es reicht, die Einträge anzugeben, die $\neq 0$ sind. Geben Sie für jedes Feld Ihren Rechenweg an.
2. Was ist die beste POS-TAG Sequenz, die sich als Ergebnis für they promise help ergibt?

Lösung:

	q_F			$108 \cdot 10^{-12}, N$	
	V		$12 \cdot 10^{-7}, \text{Pro}$	$48 \cdot 10^{-11}, V$	
	N		$1 \cdot 10^{-7}, \text{Pro}$	$108 \cdot 10^{-11}, V$	
1.	Pro	$5 \cdot 10^{-4}, q_0$			
		1 they	2 promise	3 help	

9 Pkte

2. Die beste POS-TAG Folge ist Pro V N. 1 Pkt