Sommersemester 2013 27.05.2013

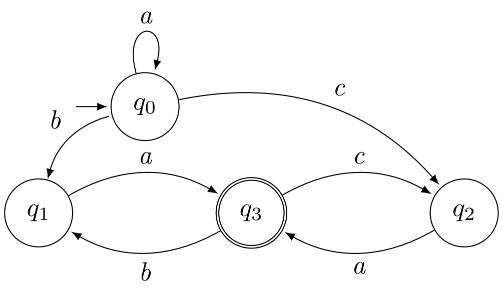
Einführung in die Computerlinguistik
Zwischenklausur

Laura Kallmeyer

Heinrich-Heine-Universität Düsseldorf

Erlaubte Hilfsmittel: Eine Din-A4 Seite mit Notizen.

Aufgabe 1 *Betrachten Sie den folgenden DFA:*



1. *Woran erkennt man, dass der Automat deterministisch ist?*
2. *Geben Sie den Automaten in Tupelschreibweise an. D.h., als Tupel* $\langle Q, \Sigma, \delta, q_0, F\rangle$.
3. *In welchem Zustand befindet sich der Automat nach den folgenden Übergängen?*
   (a) $\delta(q_1, a)$ *(b)* $\hat{\delta}(q_2, ac)$ *(c)* $\hat{\delta}(\delta(q_0, a), ba)$
4. *Geben Sie einen regulären Ausdruck für die Sprache an, die von dem Automaten akzeptiert wird.*

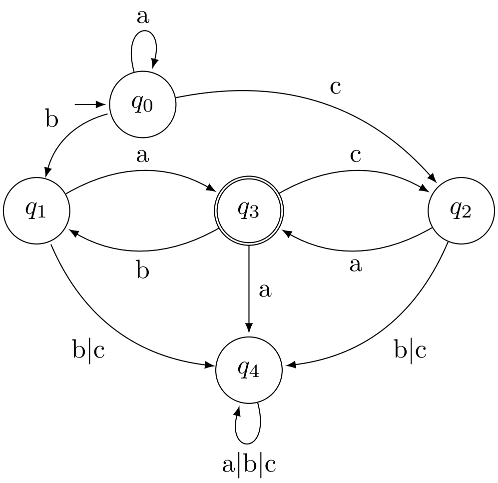
Lösung:

1. Kein Zustand hat mehr als eine ausgehende Kante mit gleichem Eingabesymbol. 1
2. $\langle\{q_0, q_1, q_2, q_3\}, \{a, b, c\}, \delta, q_0, \{q_3\}\rangle$ mit
   $\delta(q_0, a) = q_0$, $\delta(q_0, b) = q_1$, $\delta(q_0, c) = q_2$, $\delta(q_1, a) = q_3$, $\delta(q_2, a) = q_3$, $\delta(q_3, b) = q_1$, $\delta(q_3, c) = q_2$.
   Für alle anderen Paare ist δ nicht definiert. 4
3. (a) $\delta(q_1, a) = q_3$ (b) $\hat{\delta}(q_2, ac) = q_2$ (c) $\hat{\delta}(\delta(q_0, a), ba) = q_3$ 4
4. $a^*(b|c)(a(b|c))^*a$ 4

Aufgabe 2 *Betrachten Sie nochmals den DFA aus der vorhergehenden Aufgabe.*
Minimieren Sie den Automaten. Geben Sie dabei die Ergebnisse der einzelnen Schritte (Automat mit trap state, Matrix, minimierter Automat) an.

Lösung:

Automat mit Trap State:

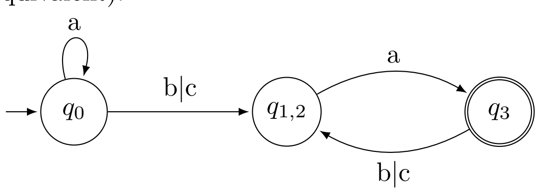


2

$|Q|x|Q|-$Matrix: 3

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

Entferne anschließend wieder den überflüssige Trap State. Minimierter Automat (q_1 und q_2 sind äquivalent):



3

Aufgabe 3 *Welche Sprachen werden von den folgenden regulären Ausdrücken denotiert?*
(a) $ab|c\emptyset$ *(b)* $a(b|c)\emptyset$ *(c)* $ab|c\varepsilon$ *(d)* $a(b|c)\varepsilon$ *(e)* $a(b^+|c^+)^*$

Lösung:

(a) $L(ab|c\emptyset) = \{ab\}$ (b) $L(a(b|c)\emptyset) = \emptyset$ (c) $L(ab|c\varepsilon) = \{ab, c\}$ (d) $L(a(b|c)\varepsilon) = \{ab, ac\}$
(e) $L(a(b^+|c^+)^*) = \{aw \mid w$ ist eine beliebige (eventuell leere) Folge von bs und cs$\}$ 6

Aufgabe 4 *Betrachten Sie folgende Grammatik:* $G = \langle\{S, A\}, \{a, b, c\}, P, S\rangle$ *mit folgenden Produktionen in* P*:* $S \to bS, S \to cS, S \to aA, A \to a$

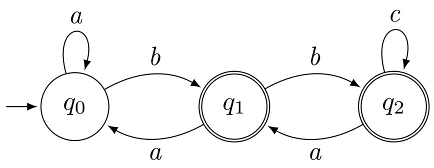
1. *Handelt es sich um eine reguläre Grammatik? Begründen Sie Ihre Antwort.*
2. *Welche Sprache wird von dieser Grammatik erzeugt?*
3. *Geben Sie eine rechtslineare Grammatik für die Sprache* $L(a^*(b|c)dd)$ *an.*

Lösung:

1. Ja, da alle rechten Seiten von Produktionen höchstens ein Nichtterminales enthalten, das immer am Ende steht. Die Grammatik ist also rechtslinear und damit regulär. 2
2. $L((b|c)^*aa)$ 3

3. Grammatik: $\langle \{S, A\}, \{a, b\}, P, S \rangle$ mit folgenden Produktionen in P:
   $S \rightarrow aS, S \rightarrow bA, S \rightarrow cA, A \rightarrow dd$ 3

Aufgabe 5 *Betrachten Sie den folgenden DFA:*



Geben Sie eine rechtslineare Grammatik an, die genau die Sprache generiert, die von diesem Automaten akzeptiert wird.

Lösung: $G = \langle \{Q_0, Q_1, Q_2\}, \{a, b, c\}, P, Q_0 \rangle$ mit folgenden Produktionen in P:

$$\begin{array}{lllll} Q_0 \rightarrow aQ_0 & Q_0 \rightarrow bQ_1 & Q_0 \rightarrow b & Q_1 \rightarrow aQ_0 & Q_1 \rightarrow bQ_2 \\ Q_1 \rightarrow b & Q_2 \rightarrow cQ_2 & Q_2 \rightarrow c & Q_2 \rightarrow aQ_1 & Q_2 \rightarrow a \end{array}$$

6

Aufgabe 6 *Nehmen Sie an, Sie haben einen HMM-POS Tagger, unter anderem mit folgenden Wahrscheinlichkeiten:*

Emissionswahrscheinlichkeiten:

$$\begin{array}{lll} P(a|Det) = 6 \cdot 10^{-3} & P(try|N) = 2 \cdot 10^{-3} & P(drink|N) = 3 \cdot 10^{-3} \\ & P(try|V) = 5 \cdot 10^{-3} & P(drink|V) = 2 \cdot 10^{-3} \end{array}$$

Alle anderen Emissionswahrscheinlichkeiten für try, a *und* drink *seien* 0.

Übergangswahrscheinlichkeiten sind (unter anderem):

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

Die Wahrscheinlichkeiten, dass der Satz mit einem V eingeleitet wird, ist $8 \cdot 10^{-1}$, *und mit einem N* $2 \cdot 10^{-1}$. *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* try a drink *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* try a drink *ergibt?*

Lösung:

1.

q_F				$432 \cdot 10^{-12}$, N
Det		$48 \cdot 10^{-7}$, V		
V	$4 \cdot 10^{-3}$, q_0		$96 \cdot 10^{-11}$, Det	
N	$4 \cdot 10^{-4}$, q_0		$432 \cdot 10^{-11}$, Det	
	1 try	2 a	3 drink	

try, V: $8 \cdot 10^{-1} \cdot 5 \cdot 10^{-3} = 40 \cdot 10^{-4} = 4 \cdot 10^{-3}$
try, N: $2 \cdot 10^{-1} \cdot 2 \cdot 10^{-3} = 4 \cdot 10^{-4}$
a, Det: $max\{4 \cdot 10^{-3} \cdot 2 \cdot 10^{-1} \cdot 6 \cdot 10^{-3}$ (Vorgänger V), $4 \cdot 10^{-4} \cdot 1 \cdot 10^{-1} \cdot 6 \cdot 10^{-3}$ (Vorgänger N)$\} = 48 \cdot 10^{-7}$ (Vorgänger V)
drink, V: $48 \cdot 10^{-7} \cdot 1 \cdot 10^{-1} \cdot 2 \cdot 10^{-3} = 96 \cdot 10^{-11}$
drink, N: $48 \cdot 10^{-7} \cdot 3 \cdot 10^{-1} \cdot 3 \cdot 10^{-3} = 432 \cdot 10^{-11}$
q_F: $max\{96 \cdot 10^{-10} \cdot 1 \cdot 10^{-1}$ (Vorgänger V), $432 \cdot 10^{-10} \cdot 1 \cdot 10^{-1}$ (Vorgänger N)$\}$ 8

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