

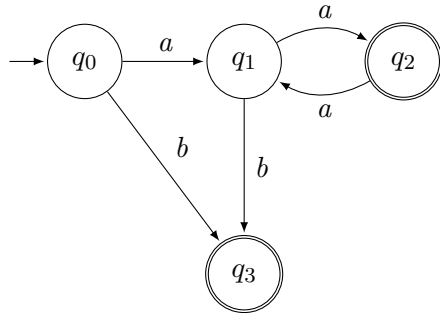
Einführung in die Computerlinguistik

Aufgaben zur Vorbereitung der Zwischenklausur

Laura Kallmeyer

SS2012, Heinrich-Heine-Universität Düsseldorf

Aufgabe 1 Gegeben sei der folgende deterministische endliche Automat:



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

Lösung:

1. $\langle Q, \Sigma, \delta, q_0, F \rangle$ mit $Q = \{q_0, q_1, q_2, q_3\}$, $\Sigma = \{a, b\}$, $F = \{q_2, q_3\}$ und δ wie folgt definiert:

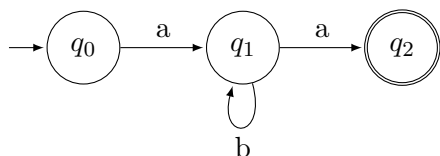
$$\delta(q_0, a) = q_1 \quad \delta(q_0, b) = q_3 \quad \delta(q_1, a) = q_2 \quad \delta(q_1, b) = q_3 \quad \delta(q_2, a) = q_1$$

2. (a) $\delta(q_1, b) = q_3$ (b) $\hat{\delta}(q_2, aa) = q_2$ (c) $\hat{\delta}(\delta(q_2, a), b) = q_3$
3. $(a(aa)^*(a|b)|b)$ (es gibt natürlich verschiedene richtige Lösungen)

Aufgabe 2 Zeichnen Sie einen deterministischen endlichen Automaten der die Sprache $L(ab^*a)$ akzeptiert und geben Sie eine rechtslineare Grammatik an, die die Sprache generiert.

Lösung:

FSA:



Grammatik: $\langle \{S, A\}, \{a, b\}, P, S \rangle$ mit folgenden Produktionen in P :

$$S \rightarrow aA \quad A \rightarrow bA \quad A \rightarrow a$$

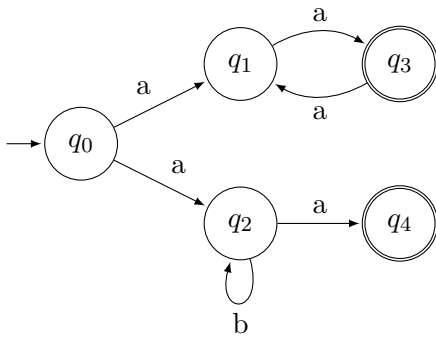
Aufgabe 3 Gegeben ist der folgende NFA $\langle Q, \Sigma, \delta, q_0, F \rangle$

- $Q = \{q_0, q_1, q_2, q_3, q_4\}$
- $\Sigma = \{a, b\}$
- $\delta(q_0, a) = \{q_1, q_2\}$
 $\delta(q_1, a) = \{q_3\}$
 $\delta(q_3, a) = \{q_1\}$
 $\delta(q_2, b) = \{q_2\} \quad \delta(q_2, a) = \{q_4\}$
- q_0 ist der Startzustand
- $F = \{q_3, q_4\}$

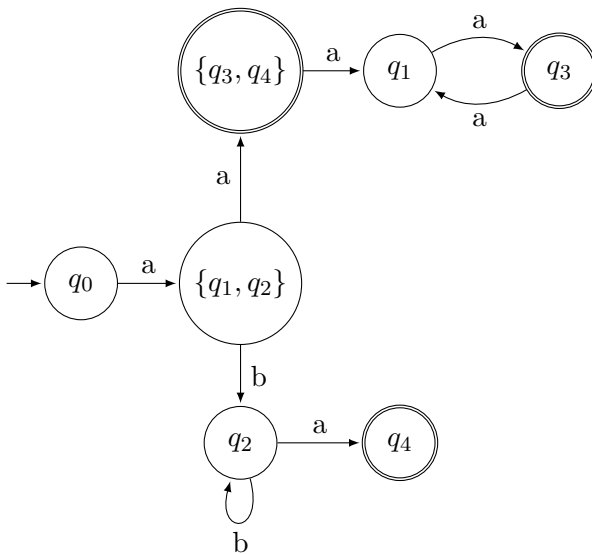
1. Woran erkennt man, dass dieser Automat nicht deterministisch ist?
2. Zeichnen Sie den NFA.
3. Konstruieren Sie einen äquivalenten DFA und geben Sie das Quintuple dazu an.
4. Geben Sie den regulären Ausdruck der Sprache an, die von dem Automaten akzeptiert wird.

Lösung:

1. Wenn von min. einem Zustand mehrere Kanten mit der gleichen Beschriftung ausgehen oder wenn er ϵ - Kanten hat.



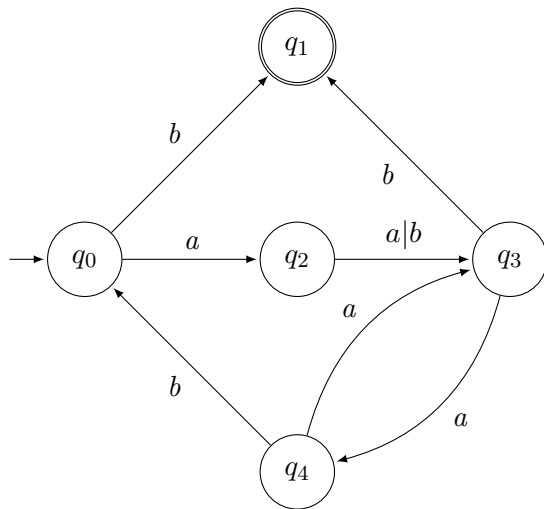
2.



3.

4. $L(R) = a(b^+|(aa)^*)a$

Aufgabe 4 Minimieren Sie den folgenden DFA.



Aufgabe 5 Geben Sie kurz wieder, wann eine CFG rechts-linear, links-linear und regulär ist.

Aufgabe 6 Gegeben sei folgende rechtslineare Grammatik:

$$S \rightarrow aB$$

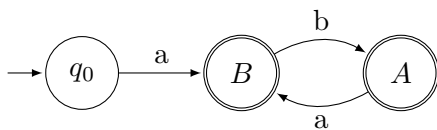
$$B \rightarrow bA|\epsilon$$

$$A \rightarrow aB|\epsilon$$

1. Geben Sie einen endlichen Automaten an, der die Sprache akzeptiert, die von der Grammatik generiert wird.
2. Wählen Sie ein Wort der Länge 5, das von der Grammatik generiert wird und zeichnen Sie den Ableitungsbaum.

Lösung:

1. Automat:



2. ...

Aufgabe 7 Nehmen Sie an, Sie haben einen HMM-POS Tagger, mit dem folgende Eingabe getaggt werden soll: learning changes thoroughly. Dem Tagger liegen folgende Wahrscheinlichkeiten zugrunde:

Emissionswahrscheinlichkeiten:

$$P(\text{learning}|V) = 3 \cdot 10^{-3} \quad P(\text{changes}|V) = 4 \cdot 10^{-3} \quad P(\text{thoroughly}|Adv) = 2 \cdot 10^{-3}$$

$$P(\text{learning}|N) = 1 \cdot 10^{-3} \quad P(\text{changes}|N) = 3 \cdot 10^{-3}$$

Alle anderen Emissionswahrscheinlichkeiten für unsere Eingabewörter seien 0.

Relevante Übergangswahrscheinlichkeiten:

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

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

Angenommen, die Wahrscheinlichkeit, dass ein N am Satzanfang steht ist $2 \cdot 10^{-1}$, die, dass ein V am Satzanfang steht $3 \cdot 10^{-1}$. Die, dass ein Satzende auf Adj folgt, ist $1 \cdot 10^{-1}$.

1. Geben Sie die Viterbi Matrix an, die sich bei diesen Wahrscheinlichkeiten für die Eingabe learning changes thoroughly ergibt. Es reicht, die Einträge anzugeben, die $\neq 0$ sind.
2. Welche POS Tags ermittelt der Tagger für die Eingabe?

Lösung:

q_F				$288 \cdot 10^{-13}$, Adv
Adv			$288 \cdot 10^{-12}$, V	
V	$9 \cdot 10^{-4}$, q_0	$36 \cdot 10^{-8}$, V		
1. N	$2 \cdot 10^{-4}$, q_0	$108 \cdot 10^{-8}$, V		
	1 learning	2 changes	3 thoroughly	

2. Die beste POS-TAG Folge ist V V Adv.

Aufgabe 8 Gegeben sei die kontextfreie Grammatik

$$G_1 = \langle \{S, A\}, \{a, b\}, S, \{S \rightarrow A|bSb, A \rightarrow aA|a\} \rangle$$

1. Welche Sprache generiert die Grammatik G_1 ?
2. Zeichnen Sie den Ableitungsbaum für das Wort bbabb.

Lösung:

1. $L = \{b^n a^m b^n \mid n \geq 0, m \geq 1\}$
2. ...

Aufgabe 9 Geben Sie für folgende Sprachen jeweils eine CFG an, die diese Sprache generiert.

1. $L_1 = \{a^n (bc)^n \mid n \geq 1\}$
2. $L_2 = \{b^n c^m d^m e^n \mid n, m \geq 0\}$

Lösung:

1. $\langle \{S\}, \{a, b, c\}, P, S \rangle$ mit folgenden Produktionen in P :

$$S \rightarrow aSbc \quad S \rightarrow abc$$

2. $\langle \{S, A\}, \{a, c, d, e\}, P, S \rangle$ mit folgenden Produktionen in P :

$$S \rightarrow bSe \quad S \rightarrow A \quad A \rightarrow cAd \quad A \rightarrow \varepsilon$$

(es gibt natürlich mehrere mögliche Lösungen)

Aufgabe 10 Betrachten Sie den folgenden Kellerautomaten M :

$M = \langle \{q\}, \{a, b\}, \{S, A, B\}, \delta, q, S, \emptyset \rangle$ mit

$$\delta(q, a, S) = \{\langle q, B \rangle\} \quad \delta(q, a, A) = \{\langle q, \varepsilon \rangle, \langle q, S \rangle\} \quad \delta(q, a, B) = \{\langle q, BB \rangle\}$$

$$\delta(q, b, S) = \{\langle q, A \rangle\} \quad \delta(q, b, B) = \{\langle q, \varepsilon \rangle, \langle q, S \rangle\} \quad \delta(q, b, A) = \{\langle q, AA \rangle\}$$

1. Geben Sie die Konfigurationen (Tripel aus Zustand, verbleibender Eingabe und Stack) an, die bei Verarbeitung von $aabb$ durchlaufen werden.
2. Welche Sprache wird von M als $N(M)$ akzeptiert? Begründen Sie Ihre Antwort, indem Sie erklären, welche Bedeutung die verschiedenen Stacksymbole haben (was man sich merkt, wenn man sie auf den Stack schiebt bzw. was man verarbeiten muss, um sie vom Stack wieder zu entfernen).

Lösung:

1. Konfigurationen:

Zust.	Resteingabe	Stack
q	$aabb$	S
q	abb	B
q	bb	BB
q	b	B
q	ε	ε

2. Ein A auf dem Stack bedeutet, dass ein Wort folgen muss, das ein a mehr als bs enthält und das zum Abbau dieses A führt,

B bedeutet, dass das Wort, bei dem das B wieder abgebaut wird, ein b mehr als as enthält.

S steht demnach für ein Wort mit gleicher Anzahl as und bs .

$$N(M) = \{w \mid w \in \{a, b\}^+, |w|_a = |w|_b\}.$$