Übungsblatt 8

Aufgabe 1. Betrachten Sie die folgende Sprache:

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

(a) Zeigen Sie, dass die Sprache L nicht regulär ist, indem sie unendlich viele Myhill-Nerode-Äquivalenzklassen angeben. Alternativ können Sie die Abschlusseigenschaften von regulären Sprachen benutzen und den Fakt, dass $L' = \{a^n b^n \mid n \geq 1\}$ nicht regulär ist.

(b) Beweisen Sie, dass L trotzdem die Eigenschaften des Pumping-Lemmas (Folie 131) erfüllt.

Lösung zu Aufgabe 1.

(a) Wir zeigen, dass unendlich viele Myhill-Nerode-Äquivalenzklassen bzgl. L existieren ($\text{index}(R_L) = \infty$) und L somit nicht regulär ist.

Behauptung: Für alle $i, k \geq 1$ mit $i \neq k$ gilt $\neg(ab^i\ R_L\ ab^k)$ und folglich $[ab^i] \neq [ab^k]$.

Sei $w = c^i$. Die Behauptung gilt, da $ab^i w = ab^i c^i \in L$ ist, während $ab^k w = ab^k c^i \notin L$ ist (wegen $i \neq k$) und somit sind beide Worte nicht Myhill-Nerode äquivalent. Es folgt, dass $[ab^i]$ für jedes $i \geq 1$ eine eigene Äquivalenzklasse ist. Folglich gilt $\text{index}(R_L) = \infty$ und L ist nicht regulär.

Alternative mit Abschlusseigenschaften: Wir definieren die Sprachen $L_1 = L \cap L(ab^*c^*) = \{ab^n c^n \mid n \geq 1\}$ und $L_2 = h(L_1)$, wobei der Homomorphismus h gegeben sei durch $h(a) = \varepsilon, h(b) = a, h(c) = b$. Somit gilt $L_2 = L'$. Wäre L regulär, so auch L_1 und folglich auch $L_2 = L'$ (wegen Abschluss regulärer Sprachen unter Schnitt- und Homomorphismenbildung). Dies ist aber ein Widerspruch, da L' nicht regulär ist. Also kann L nicht regulär sein.

(b) Wir bezeichnen den vorderen Teil von L mit A und den hinteren mit B. Sei $n = 1$. Wir zeigen: Für jedes nichtleere Wort $x \in L$ gibt es eine Zerlegung $x = uvw$ mit $|v| \geq 1, |uv| \leq 1$ und $uv^i w \in L$ für alle $i \geq 0$. Mit den Einschränkungen sehen wir bereits $u = \varepsilon$. Jetzt gibt es 3 Fälle. Falls $x = a^m b^n c^n \in A$ ist, so fängt x mit a an und wir setzen $v = a$. Dann ist $uv^i w = a^{m+i-1} b^n c^n$. Ist $m = 1$, so ist $uv^0 w = b^n c^n \in B$. Für

$m > 1$ oder $i > 0$ ist $uv^iw \in A$, da es keine Rolle spielt, wie lang der a-String vor b^nc^n ist. Im zweiten und dritten Fall ist $x = b^mc^n \in B$. Für $m > 0$ ist $v = b$ und somit $uv^iw = b^{m+i-1}c^n \in B$. Im letzten Fall ist $m = 0$ und wir haben $v = c$ und folglich $uv^iw = c^{n+i-1} \in B$.

Aufgabe 2. Sei $\Sigma = \{a, b\}$. Gegeben ist der DFA $M = (Z, \Sigma, \delta, 1, E)$ mit $Z = \{1, 2, 3, 4, 5, 6, 7\}$, $E = \{7\}$ und

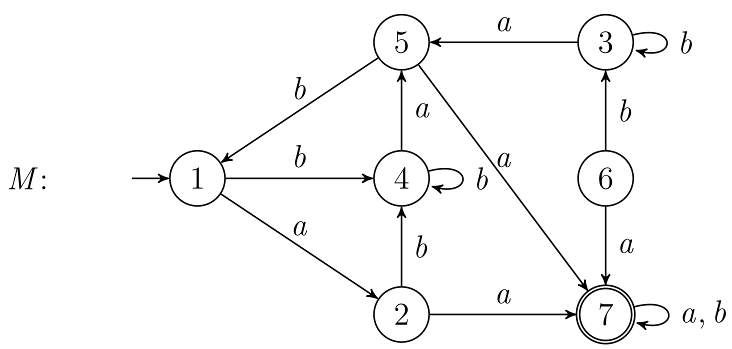
δ	a	b
1	2	4
2	7	4
3	5	3
4	5	4
5	7	1
6	7	3
7	7	7

(a) Zeichnen Sie das Automatendiagramm von M.

(b) Verwenden Sie den "Algorithmus Minimalautomat", um den Minimalautomaten für die Sprache $L(M)$ zu erhalten.

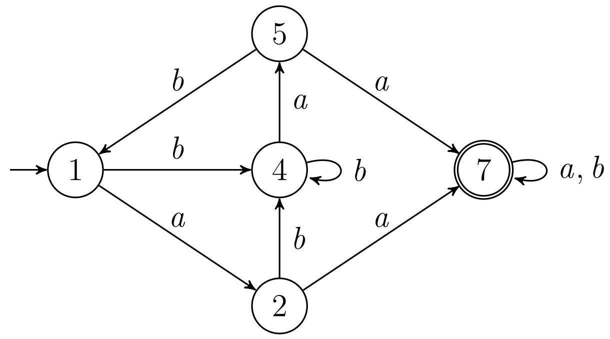
(c) Zeichnen Sie den in (b) erhaltenen Automaten.

Lösung zu Aufgabe 2. (a)



(b) Die Zustände 3 und 6 sind vom Startzustand aus nicht erreichbar und können gestrichen werden.

M:



Nun wenden wir den „Algorithmus Minimalautomat“ von Folie 165 des Skripts an.

Anmerkung: Wir arbeiten mit **Mengen** von zwei Zuständen, nicht mit Tupeln, es gilt also $\{x, y\} = \{y, x\}$.

Schritt 0

Bilden aller Zustandspaare $\{z, z'\}$ mit $z \neq z'$.

2				
4				
5				
7				
	1	2	4	5

Schritt 1

Markiere alle Paare $\{z, z'\}$ mit $z \in E$ und $z' \notin E$.

2				
4				
5				
7	1	1	1	1
	1	2	4	5

Schritt 2

Für jedes noch unmarkierte Paar $\{z, z'\}$ und jedes $s \in \Sigma$ teste, ob $\{\delta(z, s), \delta(z', s)\}$ bereits markiert ist. Falls ja, markiere auch $\{z, z'\}$.

Neue Markierungen:

- $\{1,2\}$, da $\{\delta(1,a),\delta(2,a)\} = \{2,7\}$ bereits markiert
- $\{1,5\}$, da $\{\delta(1,a),\delta(5,a)\} = \{2,7\}$ bereits markiert
- $\{2,4\}$, da $\{\delta(1,a),\delta(4,a)\} = \{7,5\}$ bereits markiert
- $\{4,5\}$, da $\{\delta(4,a),\delta(5,a)\} = \{5,7\}$ bereits markiert

	1	2	4	5
2	2			
4		2		
5	2		2	
7	1	1	1	1

Schritt 3, Wiederholung

$\{1,4\}$ und $\{2,5\}$ sind noch unmarkiert, es kommen keine weiteren Markierungen hinzu, da

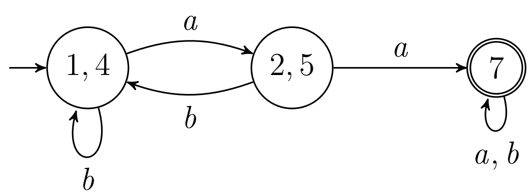
- $\{\delta(1,a),\delta(4,a)\} = \{2,5\}$ nicht markiert
- $\{\delta(1,b),\delta(4,b)\} = \{4,4\}$ nicht markiert
- $\{\delta(2,a),\delta(5,a)\} = \{7,7\}$ nicht markiert
- $\{\delta(2,b),\delta(5,b)\} = \{1,4\}$ nicht markiert

Die verbleibenden unmarkierten Zustandspaare $\{1,4\}$ und $\{2,5\}$ sind jeweils erkennungsäquivalent.

<u>Beachte:</u> Diese Begründungen müssen auch in der Klausur dazugeschrieben werden!

(c)

M:



Aufgabe 3 (Parikhs Theorem)**.** Sei $\Sigma = \{a_1, a_2, \dots, a_k\}$ ein Alphabet. Der *Parikh-Vektor* eines Wortes ist definiert als die Funktion $p : \Sigma^* \to \mathbb{N}^k$, gegeben durch

$$p(w) = (\#_{a_1}(w), \#_{a_2}(w), \dots, \#_{a_k}(w))$$

wobei $\#_{a_i}(w)$ die Anzahl der Vorkommen des Symbols a_i im Wort w bezeichnet. Eine Teilmenge von $\mathbb{N}^k$ heißt *linear*, wenn sie von der Form

$$u_0 + \mathbb{N}u_1 + \cdots + \mathbb{N}u_m = \{u_0 + t_1 u_1 + \cdots + t_m u_m \mid t_1, \dots, t_m \in \mathbb{N}\}$$

für gewisse Vektoren $u_0, \dots, u_m$ ist. Eine Teilmenge von $\mathbb{N}^k$ heißt *semilinear*, wenn sie eine endliche Vereinigung linearer Teilmengen ist.

Satz (Parikh) Sei L eine kontextfreie oder eine reguläre Sprache, und sei $P(L) = \{p(w) \mid w \in L\}$ die Menge der Parikh-Vektoren der Wörter in L (das sogenannte *Parikh-Bild* von L). Dann ist $P(L)$ eine semilineare Menge.
Ist S eine beliebige semilineare Menge, so existiert eine reguläre Sprache (und damit insbesondere auch eine kontextfreie Sprache), dessen Parikh-Bild genau S ist.

(a) Bestimmen Sie die Parikh-Bilder der folgenden Sprachen und geben Sie eine semilineare Darstellung der Menge an!

  (i) $L_1 = \{a^n b^{2n} \mid n \in \mathbb{N}\}$

  (ii) $L_2 = L((a|b|c)^* aaa(a|b|c)^*)$

  (iii) $L_3 = L(a^*|b^*)$

(b) Beweisen Sie den zweiten Teil des Satzes.

(c) Folgern Sie aus dem Satz von Parikh, dass über unären Alphabeten die Klassen der kontextfreien und regulären Sprachen gleich sind.

Lösung zu Aufgabe 3.

(a) (i) $P(L_1) = \{(x, 2x) \mid x \geq 0\} = \{t_1 \cdot (1, 2) \mid t_1 \in \mathbb{N}\}$

(ii) $P(L_2) = \{(x, y, z) \mid x \geq 3, y, z \geq 0\}$
$= \{(3, 0, 0) + t_1 \cdot (1, 0, 0) + t_2 \cdot (0, 1, 0) + t_3 \cdot (0, 0, 1) \mid t_1, t_2, t_3 \in \mathbb{N}\}$

(iii) $P(L_3) = \{(x, 0) \mid x \geq 0\} \cup \{(0, y) \mid y \geq 0\}$
$= \{t_1 \cdot (1, 0) \mid t_1 \in \mathbb{N}\} \cup \{t_1 \cdot (0, 1) \mid t_1 \in \mathbb{N}\}$

(b) Sei S eine semilineare Menge und eine semilineare Darstellung von S sei eine Vereinigung von N-vielen linearen Mengen. Ist $N = 0$, so gilt $S = \varnothing$, also ist $S = P(\varnothing)$ Parikh-Bild der regulären Sprache $\varnothing$. Für den Fall $N = 1$ (S ist eine lineare Menge) können wir S schreiben als

$$S = \{u_0 + t_1 u_1 + \cdots + t_m u_m \mid t_1, \ldots, t_m \in \mathbb{N}\}.$$

Sei $z_i \in \Sigma^*$ ein Wort mit $p(z_i) = u_i (0 \leq i \leq m)$. Betrachte die Sprache

$$L = \{z_0\} \cdot \left(\bigcup_{i=1}^{m} \{z_i\}\right)^* .$$

Dann ist $P(L) = S$ (Beweis: Übung). Außerdem ist L regulär (Abschlusseigenschaften).

Falls $N \geq 2$ gilt, so ist S endliche Vereinigung linearer Mengen, also ist S Parikh-Bild von einer Vereinigung endlich vieler regulärer Sprachen (nach dem eben Gezeigten). Diese sind aber unter Vereinigung abgeschlossen.

Insgesamt haben wir also den zweiten Teil von Parikhs Satz bewiesen.

(c) Sei L eine kontextfreie unäre Sprache. Nach dem ersten Teil von Parikhs Theorem ist $P(L)$ semilinear. Nach dem zweiten Teil gibt es eine reguläre Sprache L' mit $P(L) = P(L')$. Mit anderen Worten, die Parikh-Bilder von L und L' stimmen überein. Diese bestehen aber nur aus 1-Tupeln bzw. 1-dimensionalen Vektoren. Zu einem Parikh-Vektor (n) korrespondiert genau das Wort a^n (in L und L'), mit anderen Worten,

$$p^{-1} : \mathbb{N} \to \{a\}^*,\ n \mapsto a^n$$

ist eine Bijektion. Somit muss $L = L'$ gelten, d.h. L ist regulär.