Formale Grundlagen der Informatik I
3. Übungsblatt

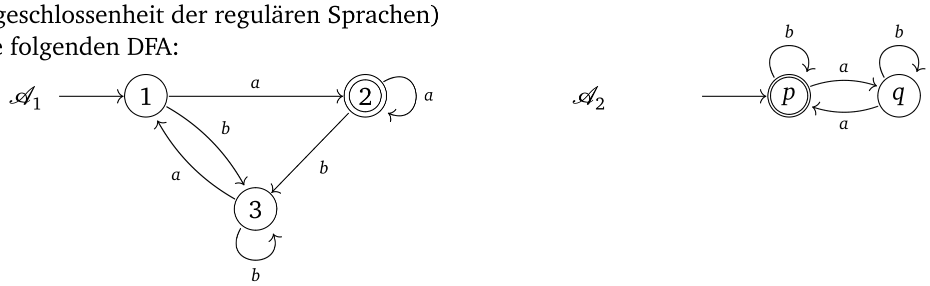
TECHNISCHE UNIVERSITÄT DARMSTADT

Fachbereich Mathematik
Prof. Dr. Ulrich Kohlenbach
Davorin Lešnik, Daniel Günzel, Daniel Körnlein

SoSe 2014
7. Mai 2014

Gruppenübung

Aufgabe G7 (Abgeschlossenheit der regulären Sprachen)
Gegeben seien die folgenden DFA:

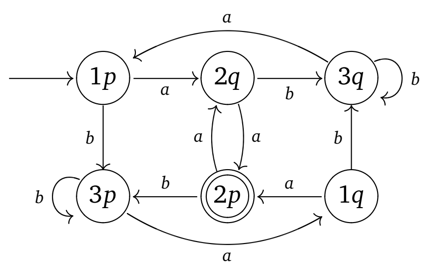


(a) Geben Sie einen DFA an, der $L(\mathscr{A}_1) \cap L(\mathscr{A}_2)$ erkennt.

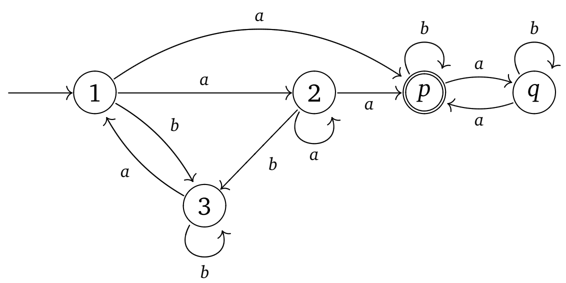
(b) Geben Sie einen NFA an, der $L(\mathscr{A}_1) \cdot L(\mathscr{A}_2)$ erkennt.
Extra: Was ändert sich an der Lösung, wenn der Zustand 1 in $\mathscr{A}_1$ auch akzeptierend ist?

Lösung:

(a) Wir bilden den Produktautomaten (vgl. Lemma 2.2.11 auf Seite 30 im Skript und Seite 72 in den Folien):



(b) Wir benutzen die Konstruktion aus Lemma 2.2.14(a) auf Seite 31 im Skript und Seite 74 in den Folien:



Falls Zustand 1 in $\mathscr{A}_1$ auch akzeptierend ist, muss in diesem Automaten der Zustand 1 auch akzeptierend sein und es muss eine a-Transition von 1 nach q sowie eine b-Transition von 1 nach p und eine a-Transition von 3 nach p hinzugefügt werden (warum?).

Aufgabe G8 (Abgeschlossenheit der regulären Sprachen)
Beweisen oder widerlegen Sie: Die Menge der regulären Sprachen ist abgeschlossen unter den folgenden Operationen:

(a) In jedem Wort werden alle Buchstaben a durch b ersetzt und alle b durch a.

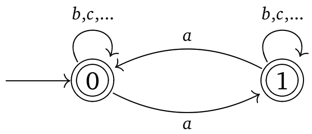
(b) Jedes zweite Vorkommen des Buchstaben a wird durch das Wort aba ersetzt.

Extra: Die Buchstaben in jedem Wort dürfen beliebig umsortiert werden, d.h. ist etwa das Wort $aaba$ in der Sprache, so fügen wir auch die Wörter $aaab$, $abaa$ und $baaa$ hinzu.

Lösung:

(a) Die Menge der regulären Sprachen ist unter dieser Operation abgeschlossen. Hat man einen regulären Ausdruck für eine Sprache, so kann man daraus einen Ausdruck für die neue Sprache konstruieren, indem man jedes a durch b und jedes b durch a ersetzt.

(b) Die Menge der regulären Sprachen ist unter dieser Operation abgeschlossen. Aus einem DFA für die alte Sprache können wir wie folgt einen Automaten für die neue Sprache konstruieren. Zuerst bilden wir das Produkt (für Durchschnitt wie in Lemma 2.2.11(a) auf Seite 30 im Skript und auf Seite 72 in den Folien) mit dem Automaten



um die Anzahl der a zu zählen. Im resultierenden Automaten können wir jede a-Transition

$$(q,1) \xrightarrow{a} (p,0),$$

welche bei einem Zustand mit zweiter Komponente 1 beginnt, ersetzen durch eine Folge von Transitionen

$$(q,1) \xrightarrow{a} \bullet \xrightarrow{b} \bullet \xrightarrow{a} (p,0).$$

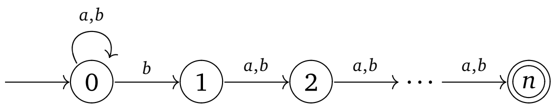
(Hierzu müssen wir für jede solche Transition zwei neue nichtakzeptierende Zwischenzustände einführen.)

Extra: Die Menge der regulären Sprache ist unter dieser Operation nicht abgeschlossen. Wenn wir diese Operation auf die sprache $L_0 = L((ab)^*)$ anwenden, erhalten wir die Sprache $L_1 = \left\{w \in \{a,b\}^* \;\middle|\; |w|_a = |w|_b\right\}$. Nach dem Schnitt mit $L(a^*b^*)$ erhalten wir

$$L_2 = L_1 \cap L(a^*b^*) = \{a^n b^n \mid n \in \mathbb{N}\}.$$

Wäre die Menge der regulären Sprachen unter der Operation abgeschlossen, dann wären auch L_1 und L_2 regulär. Aber wir wissen, dass L_2 nicht regulär ist.

Aufgabe G9 (NFA, DFA Vergleich)
Betrachten Sie den folgenden NFA $\mathscr{A}_n$:



(a) Bestimmen Sie $L(\mathscr{A}_n)$.

(b) Zeigen Sie, dass es keinen äquivalenten DFA gibt mit weniger als 2^n Zuständen.

Lösung:

(a) $L(\mathscr{A}_n)$ is die Menge von a/b-Folgen, an deren n-ter Position vor dem Ende ein b steht.

(b) Nehmen wir an, $\mathscr{Q}$ ist ein äquivalenter DFA mit weniger als 2^n Zuständen. Dann gäbe es einen Zustand q und zwei verschiedene Zeichenreihen $x_1x_2\ldots x_n$ und $y_1y_2\ldots y_n$, sodass sich $\mathscr{Q}$ nach dem Einlesen sowohl von $x_1x_2\ldots x_n$ als auch von $y_1y_2\ldots y_n$ im Zustand q befände („Schubfachprinzip").
Da die Zeichenreihen verschieden sind, müssen sie sich an einer bestimmten Position unterscheiden; sei $x_i \neq y_i$. Angenommen (auf Grund der Symmetrie ohne Beschränkung der Allgemeinheit), $x_i = b$ und $y_i = a$. Wenn $i = 1$, dann muss q sowohl ein akzeptierender Zustand als auch ein nicht akzeptierender Zustand sein, da $x_1x_2\ldots x_n$ akzeptiert wird (das n-te Zeichen vor dem Ende ist b) und $y_1y_2\ldots y_n$ nicht. Ist $i > 1$, dann betrachten wir den Zustand p, in den $\mathscr{Q}$ nach dem Einlesen sowohl von $x_1x_2\ldots x_na^{i-1}$ als auch von $y_1y_2\ldots y_na^{i-1}$ käme. Da $x_i = b$ und $y_i = a$, müsste $x_1x_2\ldots x_na^{i-1}$ akzeptiert werden (das n-te Zeichen vor dem Ende ist $x_i = b$) und $y_1y_2\ldots y_na^{i-1}$ nicht, d.h. p müsste ein akzeptierender und zugleich ein nicht akzeptierender Zustand sein.

Hausübung

Aufgabe H7 (Umkehrung regulärer Sprachen) (12 Punkte)

Für ein Wort $w = a_1 \dots a_n \in \Sigma^*$ wird w^{-1} durch $a_n \dots a_1$ definiert (d.h. w wird rückwärts gelesen). Die Sprache $\mathrm{rev}(L)$ ist definiert als

$$\mathrm{rev}(L) := \{w^{-1} \in \Sigma^* \mid w \in L\}.$$

Zeigen Sie, dass für jede reguläre Sprache L auch die Umkehrung $\mathrm{rev}(L)$ regulär ist, indem Sie zeigen, wie:

(a) man aus einem regulären Ausdruck für die Sprache L einen regulären Ausdruck für $\mathrm{rev}(L)$ gewinnen kann,

(b) aus einem NFA, der die Sprache L erkennt, ein NFA, der die Sprache $\mathrm{rev}(L)$ erkennt, allgemein konstruiert werden kann.

Lösung:

(a) Man kann induktiv über den Aufbau regulärer Ausdrücke $\alpha \in \mathrm{REG}(\Sigma)$ einen neuen regulären Ausdruck $\mathrm{rv}(\alpha)$ konstruieren, wie folgt:

(i) $\mathrm{rv}(\emptyset) := \emptyset$,

(ii) $\mathrm{rv}(a) := a$ für jedes $a \in \Sigma$,

(iii) $\mathrm{rv}(\alpha + \beta) := \mathrm{rv}(\alpha) + \mathrm{rv}(\beta)$,

(iv) $\mathrm{rv}(\alpha\beta) := \mathrm{rv}(\beta)\mathrm{rv}(\alpha)$,

(v) $\mathrm{rv}(\alpha^*) := (\mathrm{rv}(\alpha))^*$.

Wir zeigen per Induktion dann, dass für alle $\alpha \in \mathrm{REG}(\Sigma)$ gilt

$$L(\mathrm{rv}(\alpha)) = \mathrm{rev}(L(\alpha)). \tag{1}$$

Für den Induktionsanfang muss gezeigt werden, dass (1) für $\alpha = \emptyset$ und $\alpha = a$ für $a \in \Sigma$ gilt. Dies folgt sofort aus der Definition von rv.

Für die Induktionsschritt muss gezeigt werden, dass falls (1) für $\beta, \gamma \in \mathrm{REG}(\Sigma)$ eingesetzt für α gilt, dann auch (1) für $\alpha = \beta\gamma$, $\alpha = \beta + \gamma$ und $\alpha = \beta^*$ gilt. Wir zeigen hier nur den Fall $\alpha = \beta\gamma$: Gelte also

$$L(\mathrm{rv}(\beta)) = \mathrm{rev}(L(\beta)) \quad \text{und} \quad L(\mathrm{rv}(\gamma)) = \mathrm{rev}(L(\gamma)), \tag{2}$$

dann gilt

$$\begin{aligned}
L(\mathrm{rv}(\beta\gamma)) &= L(\mathrm{rv}(\gamma)\mathrm{rv}(\beta)) && \text{nach Definition von rv} \\
&= L(\mathrm{rv}(\gamma)) \cdot L(\mathrm{rv}(\beta)) \\
&= \mathrm{rev}(L(\gamma)) \cdot \mathrm{rev}(L(\beta)) && \text{nach (2)} \\
&= \mathrm{rev}(L(\beta) \cdot L(\gamma)) && \text{nach Definition von rev} \\
&= \mathrm{rev}(L(\beta\gamma)).
\end{aligned}$$

Also gilt (1) für $\alpha = \beta\gamma$.

(b) Sei $\mathscr{A} = (\Sigma, Q, q_0, \Delta, A)$ ein NFA. Für jedes $a \in A$ definieren wir

$$\mathscr{A}_a := (\Sigma, Q, q_0, \Delta, \{a\}).$$

Das heißt: $\mathscr{A}_a$ ist wie $\mathscr{A}$, aber hat nur a als akzeptierenden Zustand. Wir haben $L(\mathscr{A}) = \bigcup_{a\in A} L(\mathscr{A}_a)$ und $\mathrm{rev}(L(\mathscr{A})) = \bigcup_{a\in A} \mathrm{rev}(L(\mathscr{A}_a))$. Weil reguläre Sprachen unter Vereinigung abgeschlossen sind, brauchen wir nur einzusehen, dass $\mathrm{rev}(L(\mathscr{A}))$ regulär ist für jeden Automaten $\mathscr{A}$ mit nur einem akzeptierenden Zustand.

Sei also $\mathscr{A} = (\Sigma, Q, q_0, \Delta, \{a\})$ ein NFA mit einem akzeptierendem Zustand, der die Sprache L erkennt. Wir bestimmen einen Automaten $\mathscr{A}^{\mathrm{rev}}$, der genau $\mathrm{rev}(L)$ erkennt:

$$\mathscr{A}^{\mathrm{rev}} = (\Sigma, Q, a, \Delta^{\mathrm{rev}}, \{q_0\}), \qquad \text{wobei} \quad (q, x, q') \in \Delta^{\mathrm{rev}} \iff (q', x, q) \in \Delta.$$

Man beweist jetzt mit Induktion über n, dass es einen Lauf von q_0 nach q_n auf w in $\mathscr{A}$ genau dann gibt, wenn es einen Lauf von q_n nach q_0 auf w^{-1} in $\mathscr{A}^{\mathrm{rev}}$ gibt. Daraus folgt dann, dass $L(\mathscr{A}^{\mathrm{rev}}) = \mathrm{rev}(L)$. Wir schließen, dass auch $\mathrm{rev}(L)$ regulär ist.

Bepunktung: beide Teilaufgaben $\boxed{\text{6 P.}}$.

Aufgabe H8 (Myhill-Nerode) (12 Punkte)
Man betrachte die Sprachen

- $L_1 := \{a^m b^n \mid m, n \in \mathbb{N}\}$,
- $L_2 := \{a^m b^n \mid m, n \in \mathbb{N},\ m \leq n\}$.

(a) Bestimmen Sie den Index von $\sim_{L_1}$ und $\sim_{L_2}$.

(b) Begründen Sie, ob L_1 und L_2 reguläre Sprachen sind. Wenn ja, geben Sie einen Automaten (DFA oder NFA), der sie erkennt.

Lösung:

(a) Nach Definition gilt

$$w \sim_{L_1} w' \iff \forall\, x \in \{a,b\}^* \ \big(wx \in L_1 \iff w'x \in L_1\big)$$

für alle $w, w' \in \{a,b\}^*$. Alle Wörter, die nicht in L_1 sind, sind offensichtlich äquivalent: ein Wort w ist nicht in L_1 genau dann, wenn es ba als Teilwort enthält, und in diesem Fall gilt auch $wx \notin L_1$. Da L_1 abgeschlossen unter $\sim_{L_1}$ ist, folgt, dass $\overline{L_1}$ eine Äquivalenzklasse ist.

Für Wörter aus L_1 sind alle, die auf b enden, äquivalent, denn für solche Wörter w gilt $wx \in L_1$ genau dann, wenn x kein a enthält.

Für Wörter $w \in L_1$, die nicht auf b enden, sehen wir, dass $w \in L(a^*)$ gilt und $wx \in L_1$ genau dann, wenn $x \in L_1$. Das heißt, alle Wörter in L_1, die nicht auf b enden, sind äquivalent. Keine von diesen sind äquivalent mit Wörtern aus L_1, die auf b enden (Gegenbeispiel: $x = a$).

Die Äquivalenzklassen für $\sim_{L_1}$ sind somit $\overline{L_1} = L((a+b)^* ba (a+b)^*)$, $L(a^* b^+)$ und $L(a^*)$. Der Index (= die Anzahl der Äquivalenzklassen) von $\sim_{L_1}$ ist damit drei.

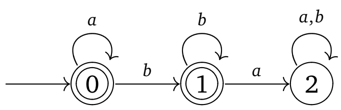
Für $\sim_{L_2}$ werden wir beweisen, dass der Index unendlich ist. Dafür genügt es, unendlich viele nicht äquivalente Wörter zu finden.

Wir zeigen, dass die Wörter $w_i := a^i$ $(i \in \mathbb{N})$, paarweise nicht äquivalent sind. Nehmen wir $i, j \in \mathbb{N}$, $i \neq j$ (o. B. d. A. $i < j$). Ein Gebenbeispiel für $a^i \sim_{L_2} a^j$ ist $x = b^i$, denn $a^i b^i \in L_2$, aber $a^j b^i \notin L_2$.

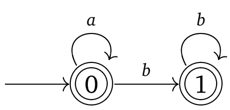
Bepunktung: $\boxed{\text{4 P.}}$ für den Index von $\sim_{L_1}$, $\boxed{\text{4 P.}}$ für den Index von $\sim_{L_2}$.

(b) $\boxed{\text{4 P.}}$ Der Satz von Myhill-Nerode besagt, dass eine Sprache genau dann regulär ist, wenn der Index ihrer Äquivalenzrelation endlich ist. Das heißt, L_1 ist regulär und L_2 nicht.

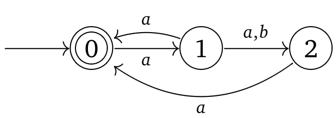
Möglicher Automat für L_1:



Bemerkung: Der angegebene DFA ist zwar minimal, denn die Anzahl der Zustände ist gleich zu dem Index. Es gibt aber noch einen kleineren NFA, nämlich



Aufgabe H9 (NFA zu DFA) (12 Punkte)
Betrachten Sie den NFA $\mathscr{A}$:

$\mathscr{A}$:


(a) Konstruieren Sie mittels Potenzmengenkonstruktion einen DFA $\mathscr{B}$, der die gleiche Sprache wie $\mathscr{A}$ erkennt.

(b) Konstruieren Sie aus $\mathscr{B}$ einen *minimalen* DFA $\mathscr{C}$, der die gleiche Sprache erkennt. Geben Sie dazu die Relationen $\nsim_i$ in tabellarischer Form an.

Lösung:

(a) $\boxed{4 \text{ P.}}$

$$\hat{A} = \Big\{\{0\}, \{0,1\}, \{0,2\}, \{0,1,2\}\Big\}$$

δ	a	b
$\{0\}$	$\{1\}$	$\emptyset$
$\{1\}$	$\{0,2\}$	$\{2\}$
$\emptyset$	$\emptyset$	$\emptyset$
$\{0,2\}$	$\{0,1\}$	$\emptyset$
$\{2\}$	$\{0\}$	$\emptyset$
$\{0,1\}$	$\{0,1,2\}$	$\{2\}$
$\{0,1,2\}$	$\{0,1,2\}$	$\{2\}$

(b) $\boxed{8 \text{ P.}}$ Wir bestimmen die Relationen $\nsim_i$ ($\times$ gehöre dabei zu $\nsim_0$ und $\times_i$ zu $\nsim_i$ für $i \geq 1$).

$\nsim$	$\{0\}$	$\{1\}$	$\emptyset$	$\{0,2\}$	$\{2\}$	$\{0,1\}$	$\{0,1,2\}$
$\{0\}$		$\times$	$\times$	$\times_1$	$\times$	$\times_1$	$\times_1$
$\{1\}$	$\times$		$\times_1$	$\times$	$\times_2$	$\times$	$\times$
$\emptyset$	$\times$	$\times_1$		$\times$	$\times_1$	$\times$	$\times$
$\{0,2\}$	$\times_1$	$\times$	$\times$		$\times$	$\times_2$	$\times_2$
$\{2\}$	$\times$	$\times_2$	$\times_1$	$\times$		$\times$	$\times$
$\{0,1\}$	$\times_1$	$\times$	$\times$	$\times_2$	$\times$		
$\{0,1,2\}$	$\times_1$	$\times$	$\times$	$\times_2$	$\times$		

Da $\nsim_3 = \nsim_2$, ist die Relation $\nsim$ durch diese Tabelle gegeben. Das heißt, dass wir die Zustände $\{1,0\}$ und $\{0,1,2\}$ identifizieren können (wir nennen den neuen Zustand X). Deshalb sieht der DFA minimaler Länge wie folgt aus: