

Einführung in die Theoretische Informatik

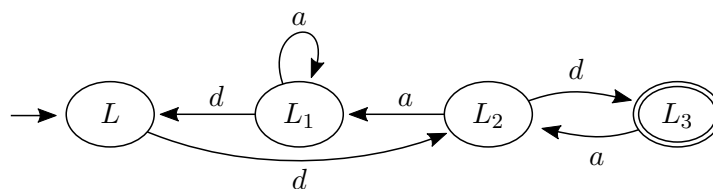
Sommersemester 2024 – Hausaufgabenblatt 5

- Die Hausaufgaben werden in folgender Reihenfolge korrigiert: **H5.4, H5.3**.
- Die Knobelaufgabe bitte separat auf Moodle abgeben. Sie wird korrigiert.

AT-Aufgabe H5.1. (*dada*) unkorrigiert (1+1+1 Punkte)

Bearbeiten Sie diese Aufgabe in [Automata Tutor](#). Der kleine Theo ist schon wieder traurig. Er würde gerne Kekse backen, muss aber zuerst herausfinden, welche Wörter sein noch kleinerer Neffe schon sagen kann. Theo hat anhand der Gesichtsausdrücke seines Neffen einen DFA erstellt, will jetzt aber wissen, welche Wörter für welche Miene möglich sind.

Abgebildet ist der kanonische Minimalautomat für $L = L(d(aa^*dd \mid da)^*d)$, fehlende Kanten führen zu einem impliziten Fangzustand. Geben Sie reguläre Ausdrücke für L_1, L_2, L_3 an. (Zur Erinnerung: Im kanonischen Minimalautomaten sind die Zustände mit Residualsprachen beschriftet.)



Lösungsskizze.

$$L_1 = L(a^*dd(aa^*dd \mid da)^*d)$$

$$L_2 = L((aa^*dd \mid da)^*d)$$

$$L_3 = L(\epsilon \mid a(aa^*dd \mid da)^*d)$$

AT-Aufgabe H5.2. unkorrigiert (2+2 Punkte)

Bearbeiten Sie diese Aufgabe in [Automata Tutor](#). Konstruieren Sie eine kontextfreie Grammatik für die folgenden Sprachen:

(a) $\{a^i b^j a^j b^i : i, j \geq 0\}$

(b) $\{a^i b^j c^k : j \geq i + k\}$

Lösungsskizze.

(a) $S \rightarrow aSb \mid X$
 $X \rightarrow bXa \mid \epsilon$

(b) $S \rightarrow XY$
 $X \rightarrow aXb \mid Xb \mid \epsilon$
 $Y \rightarrow bYc \mid \epsilon$

Aufgabe H5.3. (*DER SATZ!*) 3 + 3 + 3 Punkte

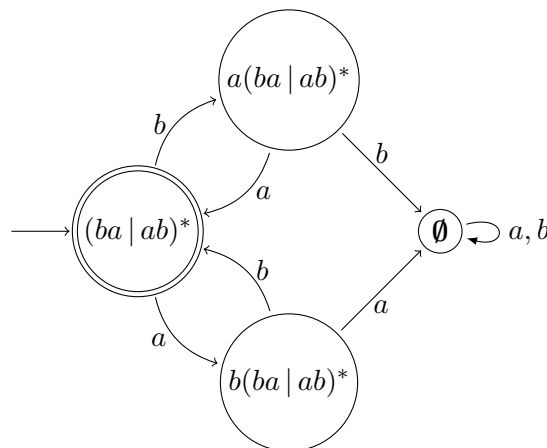
Führen Sie für jede dieser Sprachen Folgendes durch:

Falls die Sprache regulär ist, zeichnen Sie den kanonischen Minimalautomaten und beschriften Sie die Zustände mit regulären Ausdrücken für die entsprechenden Residualsprachen. Falls die Sprache nicht regulär ist, reicht es eine unendliche Menge von Wörtern mit unterschiedlichen Residualsprachen zu bestimmen und zu zeigen, dass diese Residualsprachen paarweise verschieden sind.

- (a) $L_1 := L((ba|ab)^*)$ mit dem Alphabet $\Sigma := \{a, b\}$
- (b) $L_2 := \{a^{2^k} \mid k \in \mathbb{N}\}$ mit dem Alphabet $\Sigma := \{a\}$
- (c) $L_3 := \{w \in \Sigma^* \mid |w|_{ab} \neq |w|_{ba}\}$ mit dem Alphabet $\Sigma := \{a, b\}$, wobei $|w|_v$ die Anzahl der Vorkommnisse des Wortes v im Wort w zählt. Beispielsweise gilt $|ababb|_{ab} = 2$ und $|ababb|_{ba} = 1$.

Lösungsskizze.

- (a) Kanonischer DFA:

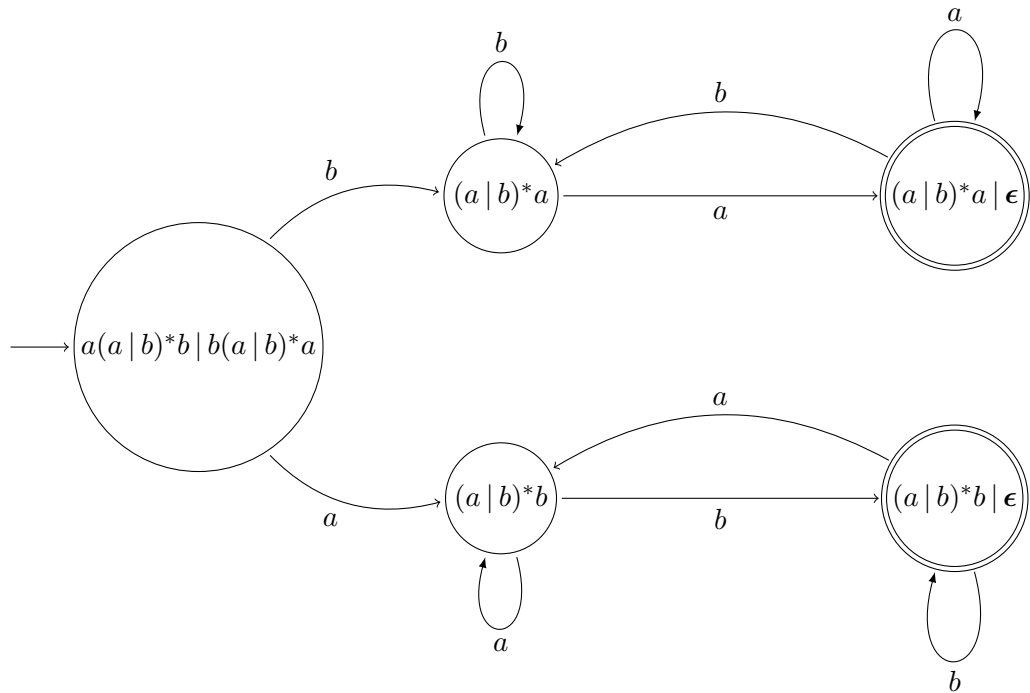


- (b) Bestimmen einer unendlichen Menge von Wörtern mit unterschiedlichen Residualsprachen:

$$\{a^{2^i} \mid i \in \mathbb{N}\}$$

Beweis: Sei $i, j \in \mathbb{N}$ verschieden und ohne Einschränkung der Allgemeinheit $i < j$. Dann gilt $a^{2^i} a^{2^i} = a^{2^{i+1}} \in L_2$. Das Wort $w = a^{2^j} a^{2^i}$ hat die Länge $2^i + 2^j = 2^i + (2^{j-i} 2^i) = 2^i (2^{j-i} + 1)$. Damit hat $|w|$ den ungeraden Teiler $2^{j-i} + 1$ und ist daher keine Zweierpotenz. Es folgt $w \notin L_2$. Somit $L^{a^{2^i}} \neq L^{a^{2^j}}$ und L_2 ist keine reguläre Sprache.

- (c) Kanonischer DFA:



Aufgabe H5.4. (*The signs were there.*)

4+2+3+1 Punkte

Der Superschurke Dr. Evilparza hat beschlossen, seinen Lebensstil grundlegend zu überdenken. Die ständigen Versuche, eine Zombiarmee aufzubauen oder einen Superlasersatelliten zu starten, sind einfach nicht gut für seine Gesundheit. Deswegen wendet er sich nun der Finanzkriminalität zu.

Sei $\Sigma := \{+, -\}$. Wir betrachten ein Bankkonto, dessen Guthaben über eine Transaktion um 1 erhöht oder gesenkt werden kann. Eine Folge von Transaktionen wird dargestellt über ein Wort in Σ^* . Wir wollen nun überprüfen, dass das Konto zu keinem Zeitpunkt überzogen wurde. Wir definieren dazu den Effekt einer Transaktion $w \in \Sigma^*$ als $\Delta(w) := |w|_+ - |w|_-$ und nennen w *überziehend*, wenn es ein i gibt, sodass $\Delta(w_1 \dots w_i) < 0$. Wir definieren L als die Sprache der Wörter, die *nicht* überziehend sind. Es gilt also z.B. $\epsilon, ++++, +-+- \in L$ und $-+, +-+--- \notin L$.

Sei $G = (\{S\}, \{+, -\}, P, S)$ die Grammatik mit den Produktionen $S \rightarrow +S-S \mid +S \mid \epsilon$.

- Zeigen Sie $L(G) \subseteq L$ mit struktureller Induktion.
- Zeigen Sie, dass jedes Wort $w \in L \setminus \{\epsilon\}$ sich in $w = +u-v$ oder $w = +u$ zerlegen lässt, mit $u, v \in L$.
- Verwenden Sie (b) und beweisen Sie $L(G) \supseteq L$.
- Ist G mehrdeutig? Begründen Sie Ihre Antwort.

Lösungsskizze. (a) Wir zeigen die Aussage über strukturelle Induktion, für die Basis überprüfen wir also, dass ϵ nicht überziehend ist und somit $\epsilon \in L$. Für den Induktionsschritt seien $u, v \in L$. Es gibt nun zwei Fälle.

- $w = +u$: Für beliebiges i gilt $\Delta(w_1 \dots w_i) = 1 + \Delta(u_1 \dots u_{i-1})$. Da u nicht überziehend ist, folgt $\Delta(u_1 \dots u_{i-1}) \geq 0$ und somit $\Delta(w_1 \dots w_i) \geq 1 \geq 0$. Also ist w auch nicht überziehend.

- $w = +u-v$: Für $i \in \{1, \dots, |u| + 2\}$ wissen wir bereits $\Delta(w_1 \dots w_i) \geq 0$, analog zum ersten Fall. Für $i > |u| + 2$ gilt nun $\Delta(w_1 \dots w_i) = \Delta(+u-v_1 \dots v_j)$, mit $j := i - |u| - 2$. Es folgt $\Delta(+u-v_1 \dots v_j) = \Delta(u) + \Delta(v_1 \dots v_j) \geq 0$, da sowohl u als auch v nicht überziehend sind.

(b) Es muss $w_1 = +$ gelten (sonst wäre $\Delta(w_1) < 0$ und $w \notin L$). Es gibt also ein x mit $w = +x$. Falls $x \in L$, sind wir bereits fertig. Ansonsten gibt es ein i mit $\Delta(x_1 \dots x_i) < 0$, wir wählen das kleinste solcher i .

Wenn $x_1 \dots x_{i-1}$ überziehend wäre, dann wäre i nicht minimal. Wir erhalten also $x_1 \dots x_{i-1} \in L$. Nun gilt $\Delta(x_1 \dots x_i) = \Delta(x_1 \dots x_{i-1}) + \Delta(x_i)$. Wir wissen, dass $\Delta(x_1 \dots x_i)$ negativ ist (nach Wahl von i), und dass $\Delta(x_1 \dots x_{i-1}) \geq 0$ (da $x_1 \dots x_{i-1}$ nicht überziehend ist). Eine einzelne Transaktion kann den Wert aber nur um $\Delta(x_i) \in \{1, -1\}$ verändern, also folgt $\Delta(x_1 \dots x_{i-1}) = 0$ und $x_i = -$.

Wir wählen unsere Zerlegung also als $u := x_1 \dots x_{i-1}$, $v := x_{i+1} \dots x_{|w|-1}$. Wir haben bereits festgestellt, dass u nicht überziehend ist, es verbleibt also noch, $v \in L$ zu zeigen.

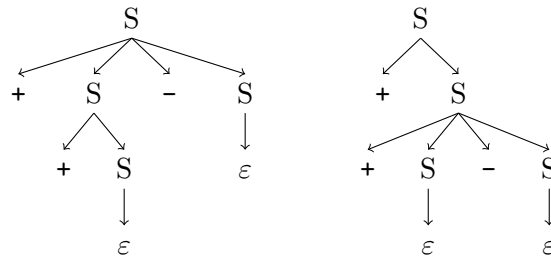
Hierfür können wir $\Delta(u) = 0$ benutzen. Sei j beliebig. Dann gilt $\Delta(v_1 \dots v_j) = \Delta(+u-) + \Delta(v_1 \dots v_j) = \Delta(w_1 \dots w_{|u|+2+j})$. Dies ist nicht negativ, da w nicht überziehend ist.

(c) Wir benutzen Induktion über die Länge des Wortes $w \in L$. Für $|w| = 0$ gilt $w \in L(G)$. Ansonsten verwenden wir die (b) und erhalten eine Zerlegung von w .

Falls $w = +u$ für ein $u \in L$, dann wissen wir nach Induktionsannahme, dass $u \in L(G)$. Daraus folgt $S \rightarrow +S \rightarrow^* +u = w$, und somit $w \in L(G)$.

Falls $w = +u-v$ für $u, v \in L$, gehen wir ähnlich vor. Nach Induktionsannahme gilt $u, v \in L(G)$, also erhalten wir $S \rightarrow +S-S \rightarrow^* +u-S \rightarrow^* +u-v = w$, und somit $w \in L(G)$.

(d) Die Grammatik ist mehrdeutig, denn das Wort $w = ++-$ hat zwei Syntaxbäume.



Quizaufgabe H5.5. (Gib mir ein a!)

unkorrigiert (3 Punkte)

Sei $G = (\{S, A, B, C\}, \{a, b\}, P, S)$ eine Grammatik mit den Produktionen

$$S \rightarrow AB \mid BC$$

$$A \rightarrow AA \mid a$$

$$B \rightarrow ABA \mid BCB \mid b$$

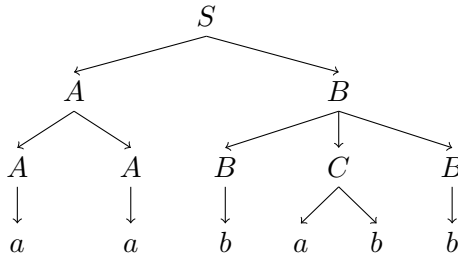
$$C \rightarrow ab \mid ba$$

Sind die folgenden Wörter in $L(G)$ enthalten? Falls ja, geben sie einen Syntaxbaum und eine Linksableitung an, falls nein, begründen Sie dies *kurz*.

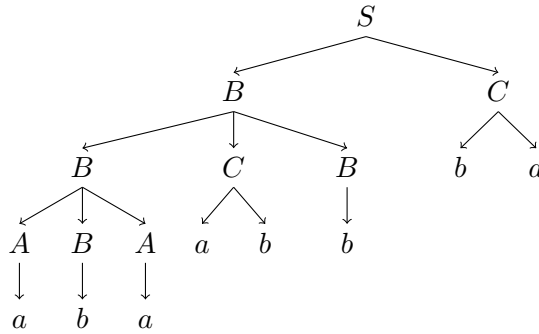
- (a) ε (b) $aababb$ (c) $abaabbba$ (d) $abaabbbba$

Lösungsskizze.

- (a) Nein, da keine Produktion $\alpha \rightarrow \varepsilon$ enthalten ist.
 (b) Ja, $S \rightarrow AB \rightarrow AAB \rightarrow aAB \rightarrow aaB \rightarrow aaBCB \rightarrow aabCB \rightarrow aababB \rightarrow aababb$.



- (c) Ja, $S \rightarrow BC \rightarrow BCBC \rightarrow ABACBC \rightarrow aBACBC \rightarrow abACBC \rightarrow abaCBC \rightarrow abaabBC \rightarrow abaabbC \rightarrow abaabbba$.



- (d) Nein, da kein Wort in G vier konsekutive b enthält. (Aus A werden nur Wörter in a^+ erzeugt, aus C nur Wörter, die ein a beinhalten und mit höchstens einem b anfangen oder aufhören. Somit aus B nur Wörter, die ein a beinhalten und mit höchstens zwei b anfangen oder aufhören, und höchstens drei konsekutive b beinhalten; ebenso für S .)

Quizaufgabe H5.6. (*Do you even pump?*)

unkorrigiert (5 Punkte)

Sei $\Sigma := \{a, b\}$. Beweisen oder widerlegen Sie, dass die folgenden Sprachen regulär sind. Bitte beachten Sie, dass es beim Beweisen von Regularität regelmäßig nicht genügt, einen entsprechenden Automaten anzugeben – zumindest eine Begründung ist erforderlich. Um zu widerlegen, dass eine Sprache regulär ist, steht Ihnen die Wahl der Beweistechnik (Pumping Lemma, unendlich viele Residualsprachen) frei.

Wie üblich schreiben wir $|x|_v := |\{u : x = uvw, u, w \in \Sigma^*\}|$ für die Anzahl der Vorkommnisse des Wortes $v \in \Sigma^*$ in $x \in \Sigma^*$, z.B. gilt $|abbbabb|_{bb} = 3$.

- (a) $L = \{w \in \Sigma^* : |w|_{aa} = |w|_{bb}\}$
- (b) $L = \{w \in \Sigma^* : |w|_{ab} = |w|_{ba}\}$

Lösungsskizze. **(a)** Nein, die Sprache ist nicht regulär. Wie zeigen dies, indem wir unendlich viele Residualsprachen finden. Wir wählen $u_i := (aab)^i$, $v_i := (abb)^i$ für $i \in \mathbb{N}$. Beachten Sie, dass $|u_i v_j|_{aa} = i$ und $|u_i v_j|_{bb} = j$ für alle $i, j \in \mathbb{N}$. In der Residualsprache L^{u_i} ist v_i enthalten, aber kein v_j für $j \neq i$. Also sind die Residualsprachen $L^{u_0}, L^{u_1}, L^{u_2}, \dots$ paarweise verschieden, und L somit nicht regulär.

(b) Ja, die Sprache ist regulär, sie enthält genau die Wörter, deren erster und letzter Buchstabe gleich sind (und das leere Wort). Dies ist dadurch begründet, dass auf ein ab kein weiteres ab folgen kann, ohne dass ein ba dazwischen entsteht, und umgekehrt. Die Differenz $|w|_{ab} - |w|_{ba}$ kann also nicht betragsmäßig größer als 1 werden. Ein regulärer Ausdruck für die Sprache ist $a(a|b)^*a | b(a|b)^*b | a | b | \epsilon$.

Knobelaufgabe H5.7. (*Irrational irregulär*)

Sei $\Sigma := \{0, \dots, 9\}$. Für Wörter $u, v \in \Sigma^*$ bezeichnen wir mit $(u.v)_{10} := (u)_{10} + 10^{-|v|} \cdot (v)_{10}$ den Wert eines Dezimalbruches, es gilt also z.B. $(1.75)_{10} = \frac{7}{4}$.

Sei $\gamma \in \mathbb{R}$, $0 \leq \gamma < 1$. Zeigen Sie, dass $\{w \in \Sigma^* : (0.w)_{10} \leq \gamma\}$ genau dann regulär ist, wenn γ rational ist.

Lösungsskizze. Wir definieren $L(x) := \{w \in \Sigma^* : (0.w)_{10} \leq x\}$ für $x \in \mathbb{R}$. Für $c \in \Sigma$ gilt

$$L(x)^c = \{w : (0.cw)_{10} \leq x\} = \{w : (0.w)_{10} \leq 10x - c\} = L(10x - c)$$

Analog folgt $L(x)^w = L(10^{|w|}x - (w)_{10})$ für $w \in \Sigma^*$.

„ \Leftarrow “: Da $\gamma \in \mathbb{Q}$ gilt, gibt es $k, n \in \mathbb{N}$ mit $\gamma = \frac{k}{n}$. Wir wollen nun zeigen, dass $L(\gamma)$ endlich viele Residualsprachen hat. Insbesondere behaupten wir $\{L(\gamma)^w : w \in \Sigma^*\} \subseteq \{L(\frac{i}{n}) : i \in \{0, \dots, n\}\}$. Diese Behauptung zeigen wir mit Induktion über die Länge von w . Die Basis folgt direkt aus $L(\gamma) = L(\frac{k}{n})$ (man bemerke $0 \leq \gamma \leq 1$). Für den Schritt sei $w \in \Sigma^*$ und $c \in \Sigma$ beliebig. Nach Annahme gilt $L(\gamma)^w = L(\frac{i}{n})$ für ein $i \in \mathbb{N}$.

$$L(\gamma)^{wc} = L(\frac{i}{n})^c = L(10 \cdot \frac{i}{n} - c) = L(\frac{10i - nc}{n})$$

Falls $j := 10i - nc$ negativ ist, gilt $L(\gamma)^{wc} = \emptyset = L(\frac{0}{n})$, für $j > n$ folgt $L(\gamma)^{wc} = \Sigma^* = L(\frac{n}{n})$, und sonst erhalten wir $L(\gamma)^{wc} = L(\frac{j}{n})$ mit $j \in \{0, \dots, n\}$. In allen drei Fällen ist die Aussage somit gezeigt.

„ \Rightarrow “: Wir zeigen zunächst, dass die Sprache $L(x)$ die Zahl x eindeutig festlegt (solange $0 \leq x < 1$). Seien $x, y \in \mathbb{R}$ mit $0 \leq x < y < 1$. Wir zeigen nun $L(x) \neq L(y)$. Dazu wählen wir ein hinreichend großes n , sodass $10^n(y - x) \geq 1$ und ein $k \in \mathbb{N}$ mit $x < k \cdot 10^{-n} \leq y$. Dann sei $w \in \Sigma^n$ mit $(w)_{10} = k$; es folgt $w \notin L(x)$ und $w \in L(y)$.

Wir wollen nun eine unendliche Folge $\gamma_0, \gamma_1, \dots \in \mathbb{R}$ konstruieren, sodass $0 \leq \gamma_i < 1$, $L(\gamma_i)$ eine Residualsprache von $L(\gamma)$ ist, und es ein $k \in \mathbb{Z}$ mit $\gamma_i = 10^i \gamma + k$ gibt, für alle $i \in \mathbb{N}$. Die Folge konstruieren wir induktiv, wir starten mit $\gamma_0 := \gamma$.

Sei nun i beliebig. Da $0 \leq \gamma_i < 1$, gibt es genau ein $c \in \Sigma$ mit $0 \leq 10\gamma_i - c < 1$. Mit $\gamma_{i+1} := 10\gamma_i - c$ gilt $L(\gamma_i)^c = L(\gamma_{i+1})$, also ist $L(\gamma_{i+1})$ eine Residualsprache von $L(\gamma)$. Schließlich gibt es ein k mit $\gamma_i = 10^i \gamma + k$, und somit $\gamma_{i+1} = 10\gamma_i - c = 10^{i+1} \gamma + 10k - c$.

Da $L(\gamma)$ regulär ist, gibt es nur endlich viele Residualsprachen. Also finden wir $i < j$ mit $L(\gamma_i) = L(\gamma_j)$. Wie wir oben argumentiert haben, folgt daraus $\gamma_i = \gamma_j$. Es gibt außerdem $k, l \in \mathbb{Z}$ mit $\gamma_i = 10^i \gamma + k$ und $\gamma_j = 10^j \gamma + l$. Insgesamt erhalten wir

$$\gamma = 10^{-i}(\gamma_i - k) = 10^{-i}(\gamma_j - k) = 10^{-i}(10^j \gamma + l - k) = 10^{j-i} \gamma + 10^{-i}(l - k)$$

Dies können wir nun nach γ auflösen (da $i \neq j$).

$$\gamma = \frac{10^{-i}(l - k)}{(1 - 10^{j-i})} = \frac{l - k}{10^i(1 - 10^{j-i})} \in \mathbb{Q}$$