Einführung in die theoretische Informatik – Aufgabenblatt 5

Beachten Sie: Soweit nicht explizit angegeben, sind Ergebnisse stets zu begründen!

Hausaufgaben: Abgabe bis zum 25.05.2016 (Mittwoch) um 12:00

Aufgabe 5.1 Levenshtein-Distanz 1P+3P+2P+2P

Wir konstruieren ein Programm das Wörter einer regulären Sprache erkennt. Leider kann es passieren, dass der Benutzer kleine Fehler bei der Eingabe macht. Das System soll diese Fehler erkennen und einen Korrekturvorschlag machen.

Wir verwenden hierfür die Levenshtein-Distanz (auch: Edit-Distanz):

Um ein Wort w in ein anderes Wort w' zu überführen, erlaubt man die Operationen *replace (R)* ($a_1 \dots a_{i-1}a_ia_{i+1} \dots a_l \to a_1 \dots a_{i-1}ba_{i+1} \dots a_l$), *delete (D)* ($a_1 \dots a_{i-1}a_ia_{i+1} \dots a_l \to a_1 \dots a_{i-1}\varepsilon a_{i+1} \dots a_l$) und *insert (I)* ($a_1 \dots a_{i-1}a_ia_{i+1} \dots a_l \to a_1 \dots a_{i-1}a_iba_{i+1} \dots a_l$) mit $a_i, b \in \Sigma$. Die Levenshtein-Distanz (geschrieben $\Delta(w, w')$) zwischen w und w' ist dann die minimale Anzahl von Operationen R, D, I, um w in w' umzuschreiben.

Wir schreiben $\Delta_{L,i} = \{w \in \Sigma^* \mid \exists w' \in L.\, \Delta(w', w) \leq i\}$ für die Sprache alle Wörter mit Edit-Distanz kleiner-gleich i zu L.

(a) Bestimmen Sie $\Delta(abcde, accd)$.

(b) Zeigen Sie: Wenn L regulär ist, dann ist auch $\Delta_{L,n}$ regulär. Konstruieren Sie hierzu einen ε-NFA N für $\Delta_{L,n}$ aus dem DFA für L. Sie müssen ausnahmsweise nicht die Korrektheit Ihres Verfahrens zeigen. Hinweis: Es ist hilfreich sich die Anzahl der bereits aufgetretenen Fehler in den Zuständen zu speichern.

(c) Beschreiben Sie ein Verfahren mit dem für ein Wort $w \in \Delta_{L,i}$ bestimmt werden kann, welches $w' \in L$ die kleinste Distanz zu w hat. Verwenden Sie hierzu die spezielle Struktur des konstruierten ε-NFAs N für $\Delta_{L,i}$.

(d) Sei nun $L = L((ab)^* \mid b^*)$. Konstruieren Sie den ε-NFA N für $\Delta_{L,1}$ und bestimmen Sie mithilfe von N alle Wörter w mit $\Delta(b, w) = 1$.

Aufgabe 5.2 Leere Sprachen 2P

Sie haben in der Vorlesung gesehen, wie man entscheiden kann, ob ein FA die leere Sprache akzeptiert. Wir entwickeln jetzt ein Verfahren, das direkt und ohne Umwege über FAs prüft, ob ein regulärer Ausdruck die leere Sprache beschreibt.

Definieren Sie hierzu eine rekursive Funktion $\mathsf{iszero} : RE \mapsto \mathbb{B}$ über den regulären Ausdrücken und beweisen Sie, dass für alle Ausdrücke r gilt:

$$\mathsf{iszero}\,(r) \Leftrightarrow L(r) = \emptyset$$

Aufgabe 5.3 Myhill-Nerode-Relation 1P+1P+1P+1P

Bestimmen Sie zu folgenden Sprachen die Myhill-Nerode-Relation, indem Sie geeignet parametrisiert alle Äquivalenzklassen aufzählen.

(a) $L_a = \{w \in \{a, b\}^* \mid |w|_a \equiv 0 \pmod 2\}$

(b) $L_b = \{w \in \{a, b\}^* \mid |w|_a = 2\,|w|_b\}$

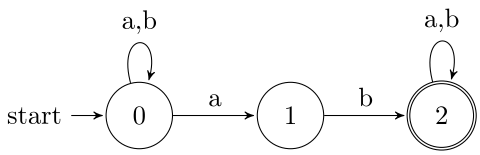
(c) $L_c = L((a^*b)^*)$

(d) $L_d = \{ww \mid w \in \{a, b\}^*\}$. Welche spezielle Eigenschaft haben die die Äquivalenzklassen: $[w]_{L_d}$?

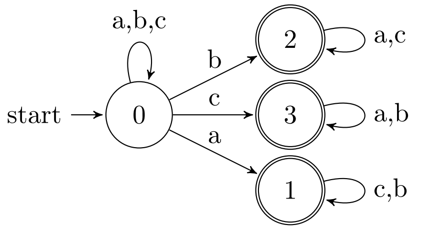
Aufgabe 5.4 **Determinisieren + Minimieren** **2P+2P+2P**

Determinisieren jeden der folgenden NFA und minimieren Sie dann jeweils den erhaltenen DFA. Konstruieren Sie bei der Determinisierung nur die vom Startzustand erreichbaren Zustände. Verwenden Sie für die Minimierung den erweiterten Algorithmus aus den Tutorübungen, um für jedes Paar von inäquivalenten Zuständen q, q' ein kürzestes Wort $w_{\{q,q'\}}$ mit $\delta(q, w_{\{q,q'\}}) \in F \leftrightarrow \delta(q', w_{\{q,q'\}}) \notin F$ zu berechnen.

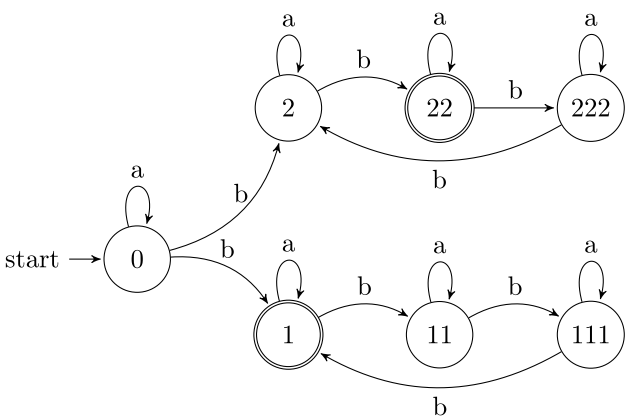
(a) NFA N_1:



(b) NFA N_2:



(c) NFA N_3:



Tutoraufgaben: Besprechung in KW20/21

Bitte beachten Sie die Regelung zu den Pfingstferien auf der Webseite.

Aufgabe 5.1

Mit $G = (\{S, E, O, A, B, X\}, \{a, b\}, P, S)$ sei die CFG mit folgenden Produktionen P bezeichnet:

$$
\begin{array}{lcl}
S & \to & E \mid O \\
E & \to & AB \mid BA \\
A & \to & XAX \mid a \\
B & \to & XBX \mid b \\
O & \to & XXO \mid X \\
X & \to & a \mid b
\end{array}
$$

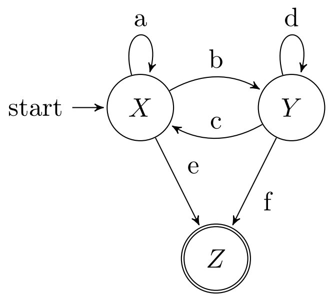
(a) Geben Sie für jedes der folgenden Wörter jeweils eine Linksableitung und eine Rechtsableitung zzgl. des entsprechenden Syntaxbaums an:

(i) *abaaaa* (ii) *babab* (iii) *aabbaaba*

(b) Bestimmen Sie die von $L(G)$ erzeugte Sprache und beweisen Sie, dass die Grammatik tatsächlich die angegebene Sprache erzeugt.

Aufgabe 5.2

Stellen Sie entsprechend der VL das zu folgendem FA gehörende Gleichungssystem auf. Bestimmen Sie dann mittels dem Gauß-Verfahren unter Verwendung von Ardens Lemma einen rationalen Ausdruck, der die vom FA akzeptierte Sprache beschreibt.



Aufgabe 5.3

Erinnerung: Ein Monoid $\langle M, \circ, 1\rangle$ besteht aus einer (Träger-)Menge M, einer assoziativen Abbildung $\circ\colon M \times M \to M$ und einem bzgl. $\circ$ neutralen Element $1 \in M$ (d.h. $\forall m \in M\colon m \circ 1 = m = 1 \circ m$). Ist $\circ$ kommutativ, dann wird $\langle M, \circ, 1\rangle$ als kommutatives Monoid bezeichnet. Gilt $\forall m \in M\colon m \circ m = m$, so wird $\langle M, \circ, 1\rangle$ als idempotentes Monoid bezeichnet.

Eine *Kleene Algebra* $\langle K, +, \cdot, ^*, 0, 1\rangle$ besteht aus einer Trägermenge K, den binären Operationen $+\colon K \times K \to K$ (Addition), $\cdot\colon K \times K \to K$ (Multiplikation), der unären Operation $^*\colon K \to K$ (Stern) und zweier nullärer Operationen/Konstanten $0, 1 \in K$. Mittels der Addition definiert man die binäre Relation $\sqsubseteq$ auf K durch

$$a \sqsubseteq b \overset{\text{Def}}{\longleftrightarrow} a + b = b$$

Wie üblich schreibt man kurz ab für $a \cdot b$. Um Klammern zu sparen, gilt "Stern vor Punkt vor Strich".

Eine Kleene Algebra erfüllt folgende Eigenschaften für alle $a, b, c, x \in K$:

Ax1: $\langle K, +, 0\rangle$ ist ein kommutatives und idempotentes Monoid.

Ax2: $\langle K, \cdot, 1\rangle$ ist ein Monoid.

Ax3: $a(b + c) = ab + ac$ und $(a + b)c = ac + bc$.

Ax4: $a0 = 0 = 0a$.

Ax5: $1 + aa^* \sqsubseteq a^*$ und $1 + a^*a \sqsubseteq a^*$.

Ax6: $b + ax \sqsubseteq x \to a^*b \sqsubseteq x$ und $b + xa \sqsubseteq x \to ba^* \sqsubseteq x$.

(a) Sei Σ ein Alphabet. Überprüfen Sie, dass $\langle 2^{\Sigma^*}, \cup, \circ, ^*, \emptyset, \{\varepsilon\}\rangle$ eine Kleene Algebra ist für

$$LL' := L \circ L' := \{ww' \mid w \in L, w' \in L'\} \qquad \text{und} \qquad L^* := \bigcup_{k \in \mathbb{N}_0} L^k$$

(b) Die übliche Addition und das Minimum auf $\mathbb{R}$ seien auf $[-\infty,\infty] = \mathbb{R} \cup \{\pm\infty\}$ wie folgt erweitert (mit $a \in \mathbb{R}$ beliebig):

$$\infty + a = \infty \qquad -\infty + a = -\infty \qquad -\infty + \infty = \infty \qquad \min(a,\infty) = a \qquad \min(a,-\infty) = -\infty \qquad \min(-\infty,\infty) = -\infty$$

Weiterhin gelte:

$$a^* := \begin{cases} -\infty & \text{falls } a \in [-\infty, 0) \\ 0 & \text{falls } a \in [0,\infty] \end{cases}$$

Überprüfen Sie wieder, dass $\langle \mathbb{R} \cup \{\pm\infty\}, \min, +, {}^*, \infty, 0\rangle$ eine Kleene Algebra ist.

(c) Zeigen Sie, dass in jeder Kleene Algebra $\langle K, +, \cdot, {}^*, 0, 1\rangle$ für beliebige $a,b,c,d,e,f \in K$ gilt:

(i) $\sqsubseteq$ ist eine partielle Ordnung auf K, die monoton bzgl. Addition, Multiplikation und Stern ist, d.h.:

$$a \sqsubseteq b \to (ca \sqsubseteq cb \wedge ac \sqsubseteq bc \wedge a + c \sqsubseteq b + c \wedge a^* \sqsubseteq b^*)$$

(ii) a^*b ist die bzgl. $\sqsubseteq$ kleinste Lösung der linearen Ungleichung $b + aX \sqsubseteq X$ in K (X Variable), genauer:

$$b + a(a^*b) = a^*b \qquad \wedge \qquad \forall x \in K\colon b + ax \sqsubseteq x \to a^*b \sqsubseteq x$$

Entsprechend ist ba^* die kleinste Lösung in K von $b + Xa \sqsubseteq X$.

(d) (i) (Knobelaufgabe) Zeigen Sie mittels Induktion nach n, dass jedes lineare Ungleichungsssystem

$$\begin{array}{rcl} a_{1,1}X_1 + a_{1,2}X_2 + \ldots a_{1,n}X_n + b_1 & \sqsubseteq & X_1 \\ \vdots & & \\ a_{n,1}X_1 + a_{n,2}X_2 + \ldots a_{n,n}X_n + b_n & \sqsubseteq & X_n \end{array}$$

mit $a_{i,j}, b_i \in K$ und $X_1,\ldots,X_n$ Variablen stets eine eindeutige $\sqsubseteq$-kleinste Lösung in K hat, d.h. dass es konkrete Elemente $x_1,\ldots,x_n \in K$ gibt, so dass für $X_1 = x_1,\ldots,X_n = x_n$ alle obigen Ungleichungen erfüllt sind, und für jede weitere Lösung $X_1 = y_1,\ldots,X_n = y_n \in K$ stets $x_i \sqsubseteq y_i$ gilt. Insbesondere gilt

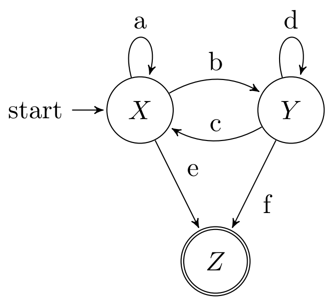
$$x_1 = a_{1,1}^*(a_{1,2}x_2 + \ldots + a_{1,n}x_n + b_1)$$

Machen Sie sich klar, dass man hiermit das Gauß-Verfahren zur Bestimmung von $x_1,\ldots,x_n$ verwenden kann.

(ii) Bestimmen Sie die kleinste Lösung x_1, x_2, x_3, x_4 für folgendes System:

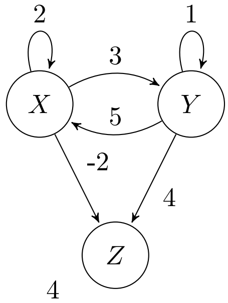
$$\begin{array}{rcl} aX + bY + eZ & \sqsubseteq & X \\ cX + dY + fZ & \sqsubseteq & Y \\ 1 & \sqsubseteq & Z \end{array}$$

(iii) Zeigen Sie, dass der für X in (5.2) berechnete Term gerade die Sprache $L(N)$ mit N der folgende NFA beschreibt, wenn man den Term über $\langle 2^{\Sigma^*}, \cup, \circ, {}^*, \emptyset, \{\varepsilon\}\rangle$ auswertet:



(iv) Bestimmen Sie die Länge eines kürzesten Pfades von jedem Knoten zum Knoten Z in folgendem gewichteten Graphen.

Durch welche Werte muss man a,b,c,d,e,f konkret ersetzen, damit sich die in (d-ii) berechneten Terme in $\langle \mathbb{R} \cup \{\pm\infty\}, \min, +, \infty, 0\rangle$ zu den gesuchten Pfadlängen auswerten?



Knobelaufgabe

Freiwillige Abgabe per E-Mail an theo-uebungsleitung@in.tum.de. Alle richtigen Einsendungen bis zum 25.05.2016 um 12:00 Uhr erhalten die Bonuspunkte. Wir bewerten nur die erste Einsendung und behalten uns das Recht vor unleserliche oder unverständliche Abgaben nicht zu korrigieren.

Aufgabe 5.1

Zeigen Sie die Behauptung aus TA5.3(c)(i)!