Prof. Dr. Jasmin Blanchette
Jannis Limperg

Ludwig-Maximilians-Universität München
Institut für Informatik
Abgabe bis 17. Mai 2023, 10:00 Uhr

Lösungsvorschlag zur Übung 4 zur Vorlesung
Theoretische Informatik für Medieninformatiker

TIMI4-1 ***Reguläre Ausdrücke*** (2 Punkte)

In dieser Aufgabe sind reguläre Ausdrücke entsprechend der Definition und Syntax im Vorlesungsskript, Definition 4.7.1, anzugeben.

a) Geben Sie einen regulären Ausdruck an, der genau die Sprache

$$L = \{w \mid i \in \Sigma, w \in \Sigma^* \text{ und } \#_i(w) = i\}$$

erzeugt, wobei $\Sigma = \{1, 2, 3\}$. (Siehe auch TIMI3-1b.)

LÖSUNGSVORSCHLAG:

$$(\underbrace{((2\,|\,3)^*1(2\,|\,3)^*)}_{\text{genau eine 1}} \;|\; \underbrace{((1\,|\,3)^*2(1\,|\,3)^*2(1\,|\,3)^*)}_{\text{genau zwei 2en}} \;|\; \underbrace{((1\,|\,2)^*3(1\,|\,2)^*3(1\,|\,2)^*3(1\,|\,2)^*)}_{\text{genau drei 3en}})$$

b) Für Autokennzeichen in München gibt es folgende Regeln:

- Sie beginnen mit einem M, danach folgt ein Buchstabenfeld, danach folgt ein Ziffernfeld. Anschließend können noch ein H für Oldtimer oder ein E für Elektroautos kommen, aber nicht beides gleichzeitig.
- Das Buchstabenfeld besteht aus 1 oder 2 Buchstaben aus $\{A, \ldots, Z\}$.
- Das Ziffernfeld besteht aus 3-4 Ziffern $\{0, \ldots, 9\}$, die erste Ziffer darf nicht 0 sein.
- Gehört das Fahrzeug allerdings einer Behörde an, so ist das Buchstabenfeld leer und der Ziffernblock 1-5 Ziffern lang. Dann darf der Ziffernblock ebenfalls nicht mit einer 0 anfangen und falls er 3 oder mehr Ziffern lang ist, auch nicht mit einer 4.
- Anmerkung: Es gibt weitere Sonderregelungen, welche in dieser Aufgabe ignoriert werden.

Geben Sie einen regulären Ausdruck für die gültigen Münchner Autokennzeichen an.

LÖSUNGSVORSCHLAG:

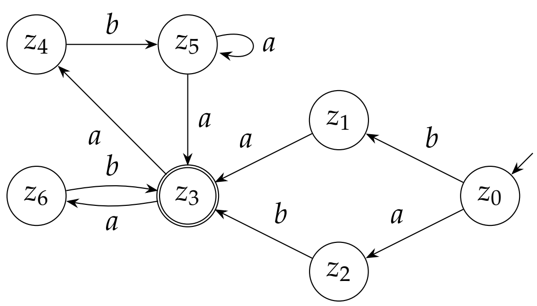
Teile des Ausdrucks:

- $X_1 := A \mid \cdots \mid Z$ sind die möglichen einbuchstabigen Autokennzeichen
- $X_2 := (A \mid \cdots \mid Z)(A \mid \cdots \mid Z)$ sind die möglichen zweibuchstabigen Autokennzeichen
- $X_3 := (1 \mid \cdots \mid 9)(0 \mid \cdots \mid 9)(0 \mid \cdots \mid 9)(0 \mid \cdots \mid 9 \mid \varepsilon)$ sind die 3-4-stelligen Zahlen
- $X_4 := (1 \mid 2 \mid 3 \mid 5 \mid \cdots \mid 9)(0 \mid \cdots \mid 9 \mid \varepsilon)(0 \mid \cdots \mid 9 \mid \varepsilon)(0 \mid \cdots \mid 9 \mid \varepsilon)(0 \mid \cdots \mid 9 \mid \varepsilon)$ sind die 1-5-stelligen Zahlen, die nicht mit einer 4 anfangen
- $X_5 := 4(0 \mid \cdots \mid 9 \mid \varepsilon)$ sind die 1-2-stelligen Zahlen, die mit einer 4 anfangen

$M((X_1 \mid X_2)X_3 \mid X_4 \mid X_5)(E \mid H \mid \varepsilon)$ ist damit der Gesamtausdruck.

Ausgeschrieben: $M(((A \mid \cdots \mid Z) \mid ((A \mid \cdots \mid Z)(A \mid \cdots \mid Z)))((1 \mid \cdots \mid 9)(0 \mid \cdots \mid 9)(0 \mid \cdots \mid 9)(0 \mid \cdots \mid 9 \mid \varepsilon)) \mid ((1 \mid 2 \mid 3 \mid 5 \mid \cdots \mid 9)(0 \mid \cdots \mid 9 \mid \varepsilon)(0 \mid \cdots \mid 9 \mid \varepsilon)(0 \mid \cdots \mid 9 \mid \varepsilon)(0 \mid \cdots \mid 9 \mid \varepsilon)) \mid (4(0 \mid \cdots \mid 9 \mid \varepsilon)))(E \mid H \mid \varepsilon)$

c) Betrachten Sie folgenden NFA A:



Geben Sie einen regulären Ausdruck an, der die von A erkannte Sprache $L(A)$ erzeugt.

Hinweis: Sie müssen nicht das Verfahren aus der Vorlesung befolgen. Ein Vorgehen ist es, die vom NFA erkannte Sprache zunächst in Wörtern zu beschreiben und sich im Anschluss daran einen regulären Ausdruck dafür zu überlegen.

LÖSUNGSVORSCHLAG:

Zunächst muss vom Startzustand z_0 der Zustand z_3 erreicht werden. Daher beginnt jedes akzeptierte Wort mit ab oder mit ba. Um von z_3 wieder in (den einzigen akzeptierenden Zustand) z_3 zu gelangen, gibt es zwei Pfade (Kreise): Entweder über z_6, dies ist mit ab möglich, oder über $z_4 \to z_5 \to z_3$, wobei man in z_5 noch beliebig lange durch Lesen von a verbleiben kann, d.h. dieser Weg entspricht den Teilwörten dargestellt durch aba^*a. Man kann über die beiden Kreise z_3 immer wieder besuchen, daher ergibt sich ingesamt der reguläre Ausdruck

$$(ab \mid ba)(ab \mid aba^*a)^*$$

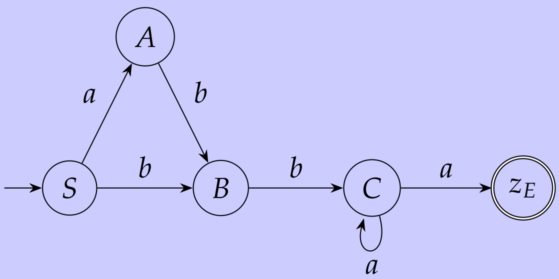
TIMI4-2 ***Grammatik über Automaten zu Grammatik*** (0 Punkte)

Gegeben sei die Grammatik

$$G = (\{S, A, B, C\}, \{a, b\}, \{S \to aA \mid bB,\ A \to bB,\ B \to bC,\ C \to aC \mid a\}, S)$$

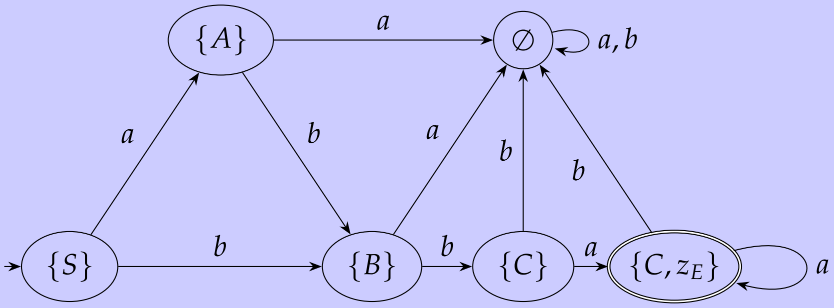
a) Erzeugen Sie gemäß der Konstruktion aus der Vorlesung aus G einen NFA A mit $L(G) = L(A)$. Zeichnen Sie den Zustandsgraph von A.

LÖSUNGSVORSCHLAG:



b) Erzeugen Sie mit der Potenzmengenkonstruktion aus A einen DFA B mit $L(B) = L(A)$. Zeichnen Sie den vom Startzustand erreichbaren Teil des Zustandsgraphen von B.

LÖSUNGSVORSCHLAG:



c) Erzeugen Sie gemäß der Konstruktion aus der Vorlesung aus B eine Grammatik H mit $L(B) = L(H)$.

LÖSUNGSVORSCHLAG:

Neubenennung der Zustände (um sie klarer erkenntlich als Variablennamen schreiben zu können):

$V_S := \{S\}, V_A := \{A\}, V_B := \{B\}, V_C := \{C\}, V_{CE} := \{C, z_E\}, V_\emptyset = \emptyset$

$H = (\{V_S, V_A, V_B, V_C, V_{CE}, V_\emptyset\}, \{a, b\}, \{$

$V_S \to aV_A \mid bV_B,$

$V_A \to aV_\emptyset \mid bV_B,$

$V_B \to aV_\emptyset \mid bV_C,$

$V_C \to aV_{CE} \mid a \mid bV_\emptyset,$

$V_{CE} \to aV_{CE} \mid a \mid bV_\emptyset,$

$V_\emptyset \to aV_\emptyset \mid bV_\emptyset,$

$\}, V_S)$

d) Vergleichen Sie die Grammatiken G und H. Beschreiben Sie die Gemeinsamkeiten dieser Grammatiken, sowie ihre Unterschiede.

Überlegen Sie sich, wodurch diese Effekte zustande kommen.

LÖSUNGSVORSCHLAG:

Grund für diese Effekte ist, dass G direkt einem nichtdeterministischen Automaten entspricht und H direkt einem deterministischen Automaten. Darum übertragen sich die Eigenschaften der entsprechenden Automaten und Automatenmodelle auf die Grammatiken.

Gemeinsamkeiten:

- Sie akzeptieren die gleiche Sprache (aufgrund der Konstruktion).
- Es gibt sowohl in G als auch in H Variablen, die ein Terminalsymbol und sich selbst erzeugen.

Unterschiede:

- H enthält mehr Variablen und deutlich mehr Regeln als G (da die Automaten unterschiedlich groß sind).
- H enthält die Variable $V_\emptyset$, über die beliebig („unendlich“) lange abgeleitet werden kann, ohne dass je ein Wort produziert würde. Dadurch kann man für jedes Teilwort $w \in \Sigma^*$ eine Satzform wV_x (wobei V_x eine Variable ist) ableiten. Wendet man die Regeln von H „blind“ an, kommt man also leicht in eine Sackgasse. G hat dieses Problem nicht.