

Einführung in die Computerlinguistik

Hausaufgabe 2, Abgabe 04.05.2020

Laura Kallmeyer

SoSe 2020, Heinrich-Heine-Universität Düsseldorf

Aufgabe 1 Geben Sie die Denotate der folgenden regulären Ausdrücke an, indem Sie die jeweilige Menge, falls sie endlich ist, explizit auflisten, und sonst in Worten beschreiben.

1. $\emptyset\varepsilon$ 2. $(a|\varepsilon|b)(ab|\varepsilon)$ 3. $\varepsilon(\varepsilon|\emptyset)$ 4. $ab^+b(c|\varepsilon)$ 5. $((a|\varepsilon)(b|c))^*$ 6. $(ac|ac\emptyset|b\emptyset|\varepsilon b|ab|c)^*$

Lösung:

1. \emptyset
2. $\{\varepsilon, a, b, ab, aab, bab\}$
3. $\{\varepsilon\}$
4. $\{ab^n \mid n \geq 2\} \cup \{ab^n c \mid n \geq 2\}$
5. $\{w \in \{a, b, c\}^* \mid \text{auf jedes } a \text{ in } w \text{ folgt notwendig als nächstes entweder ein } b \text{ oder ein } c\}$
6. gleiche Sprache wie 5.

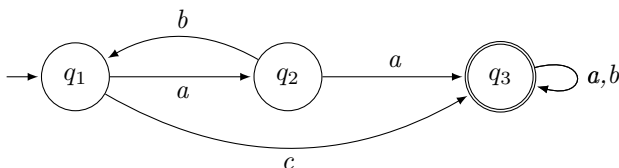
Aufgabe 2 Geben Sie für die folgenden Mengen jeweils einen regulären Ausdruck an, der die jeweilige Menge denotiert.

1. $\{w \mid w \in \{a, b, c\}^* \text{ und } w \text{ enthält entweder kein } b \text{ oder kein } c\}$
2. $\{w \mid w \in \{a, b, c\}^+, \text{ alle } a\text{'s und } b\text{'s in } w \text{ sind notwendig von einem } c \text{ gefolgt, } c\text{'s dürfen darüber hinaus beliebig auftreten}\}$
3. $\{w \mid w \in \{a, b\}^* \text{ und in } w \text{ muss nach jedem zweiten } a \text{ ein } b \text{ folgen. Was ansonsten an } b\text{'s zwischen den } a\text{'s steht, ist egal, und es darf auch gar kein } a \text{ oder nur ein } a \text{ geben}\}$
Die Sprache enthält z.B. bbba, a, bbabab, aab, abbbaba, baabbbbabbab, usw.

Lösung:

1. $(a|c)^*(a|b)^*$ 2. $(ac|bc|c)^+$ 3. $(b^*ab^*ab)^*b^*(a|\varepsilon)b^*$

Aufgabe 3 Gegeben sei der folgende DFA:



Berechnen Sie einen regulären Ausdruck, der die Sprache beschreibt, die von dem Automaten akzeptiert wird, indem Sie den Algorithmus aus der Vorlesung anwenden.

Hinweis: Wenn es keinen Pfad gibt, lautet der entsprechende reguläre Ausdruck \emptyset .

Lösung:

Wir haben einen Endzustand, der Ausdruck, den wir benötigen, ist also $r_{1,3}^3$

- $r_{1,3}^3 = r_{1,3}^2 | r_{1,3}^2 (r_{33}^2)^* r_{33}^2$ (rek. Formel)
- $r_{1,3}^2 = r_{1,3}^1 | r_{1,2}^1 (r_{22}^1)^* r_{23}^1$ (rek. Formel)
 $r_{1,3}^1 = c, r_{1,2}^1 = a, r_{22}^1 = ba|\varepsilon, r_{23}^1 = bc|a$, also $r_{1,3}^2 = c|a(ba|\varepsilon)^*(bc|a) = c|a(ba)^*(bc|a)$
- $r_{33}^2 = r_{33}^1 | r_{32}^1 (r_{22}^1)^* r_{23}^1$ (rek. Formel)
 $= (a|b|\varepsilon) | \emptyset (ba|\varepsilon)^* (bc|a) = (a|b|\varepsilon)$
- damit ergibt sich insgesamt
 $r_{1,3}^3 = (c|a(ba)^*(bc|a)) | (c|a(ba)^*(bc|a)) (a|b|\varepsilon)^* (a|b|\varepsilon) = (c|a(ba)^*(bc|a)) (a|b)^*$

Aufgabe 4 Geben Sie für die folgenden Mengen jeweils eine reguläre Grammatik an, die diese Menge generiert.

1. $L((ab)^+cb^+)$
2. $\{w \mid w \in \{a,b\}^* \text{ und die Länge von } w \text{ ist gerade (0 ist auch erlaubt)}\}$

Lösung:

1. $G = \langle N, T, P, S \rangle$ mit $N = \{A, B, S\}, T = \{a, b, c\}$ und
 $P = \{S \rightarrow abA, A \rightarrow abA, A \rightarrow cB, B \rightarrow bB, B \rightarrow b\}$
2. $G = \langle N, T, P, S \rangle$ mit $N = \{A, S\}, T = \{a, b\}$ und
 $P = \{S \rightarrow aA, S \rightarrow bA, A \rightarrow aS, A \rightarrow bS, S \rightarrow \varepsilon\}$