Thomas Schwentick

technische universität dortmund

Gaetano Geck, Lutz Oettershagen,
Christopher Spinrath, Marco Wilhelm

Lehrstuhl 1
Logik in der Informatik

SoSe 2018 | **Übungsblatt 1** | **12.04.2018**

Aufgabe 1.1 [Reguläre Ausdrücke vergleichen und interpretieren] — 3 Punkte

a) Seien $\beta = (ab)^*$ und $\alpha_1, \ldots, \alpha_8$ die folgenden erweiterten regulären Ausdrücke. Beurteilen Sie für alle $i \in \{3, \ldots, 8\}$, ob $L(\alpha_i) \subseteq L(\beta)$ gilt. Vervollständigen Sie dazu die folgende Tabelle analog zu den Beispiel-Ausdrücken α_1, α_2: Falls $L(\alpha_i) \nsubseteq L(\beta)$ gilt, geben Sie ein Wort $w_i \in L(\alpha_i) - L(\beta)$ an. **(2 Punkte)**

RE	$L(\alpha_i) \subseteq L(\beta)$	Gegenbeispiel
$\alpha_1 = (ab)^*(ab)^*$	✓	
$\alpha_2 = (ba)^*$	✗	ba
$\alpha_3 = (a^*b^*)$		
$\alpha_4 = (b?a?)^*$		
$\alpha_5 = (a+\varepsilon)(b+\varepsilon)(ab)^*$		
$\alpha_6 = (ab^*ab^*)^*$		
$\alpha_7 = (abab)^*$		
$\alpha_8 = a(ba)^+b$		

b) Das Verhalten eines Netzwerk-Controllers soll anhand protokollierter Ausgaben analysiert werden. Der Controller schreibt, abhängig von der Eingabe, beliebig lange Bitfolgen. Eine Bitfolge wird als Wort über dem Alphabet $\{0, 1\}$ repräsentiert. Die Menge der gültigen Ausgaben wird im Handbuch des Controllers formal als Sprache $L = \{0, 1\}^* - L(\gamma)$ für den erweiterten regulären Ausdruck

$$\gamma = (0+1)^*(000+111)(0+1)^* + (10)^*1? + (01)^*0?$$

spezifiziert. Beschreiben Sie L natürlichsprachlich kurz in einem Satz. **(1 Punkt)**

Lösung:

a) Eine korrekte Vervollständigung der Tabelle (mit gleich mehreren Gegenbeispielen) ist nachfolgend angegeben.

RE	$L(\alpha_i) \subseteq L(\beta)$	Gegenbeispiel
$\alpha_1 = (ab)^*(ab)^*$	✓	
$\alpha_2 = (ba)^*$	✗	ba
$\alpha_3 = (a^*b^*)$	✗	$a, b, aa, bb, ba, \ldots$
$\alpha_4 = (b?a?)^*$	✗	$a, b, aa, bb, ba, \ldots$
$\alpha_5 = (a+\varepsilon)(b+\varepsilon)(ab)^*$	✗	$a, b, aab, bab, \ldots$
$\alpha_6 = (ab^*ab^*)^*$	✗	$aa, abba, \ldots$
$\alpha_7 = (abab)^*$	✓	
$\alpha_8 = a(ba)^+b$	✓	

b) Die Sprache L enthält genau die Wörter, die weder dreimal dasselbe Zeichen aufeinanderfolgend enthalten (000 oder 111) noch eine Folge wechselnder Bits sind.

Dies folgt daraus, dass L das Komplement der von γ beschriebenen Sprache $L(\gamma)$ ist und es gilt

$$L(\gamma) = \underbrace{L\big((0+1)^*(000+111)(0+1)^*\big)}_{\text{Teilwort 000 oder 111 enthalten}} \cup \underbrace{L\big((10)^*1?\big)}_{\text{wechs. Bits, Anfang mit 1}} \cup \underbrace{L\big((01)^*0?\big)}_{\text{wechs. Bits, Anfang mit 0}}.$$

Der Ausdruck $(0+1)^*(000+111)(0+1)^*$ erzeugt alle Wörter, die 000 oder 111, möglicherweise sogar beide, enthalten: Ihnen können vermöge der Kombination von Auswahl- und Wiederholungsoperator, $(0+1)^*$, beliebig viele Bits voran- und nachgestellt werden. Der Ausdruck $(10)^*1?$ erzeugt vermöge der Konkatenation alle Wörter der Form $w = w'w''$, wobei w' durch $(10)^*$ erzeugt wird, also von der Form $(10)^n$ für $n \in \mathbb{N}_0$ ist, und $w'' \in \{1, \varepsilon\}$ ist. Insbesondere werden die Wörter $\varepsilon, 1, 10, 101, 1010, 10101$ erzeugt. Der Ausdruck $(01)^*0?$ ist analog konzipiert (mit invertierten Bits).

Aufgabe 1.2 [Reguläre Ausdrücke konstruieren] **6 Punkte**

Geben Sie im Folgenden reguläre Ausdrücke bzw. erweiterte reguläre Ausdrücke an. Beschreiben Sie für jede Konstruktion kurz, warum Ihr Ausdruck die Sprache beschreibt (warum er *alle* Wörter der Sprache erzeugt und warum er *kein* Wort außerhalb der Sprache erzeugt).

Hinweis

Sie dürfen in dieser Aufgabe zunächst Teilausdrücke konstruieren und diese dann zu größeren Ausdrücken kombinieren. Beispielsweise dürften Sie

$$\begin{aligned} \alpha &= (a+b)(a+b) \\ \beta &= (\alpha\alpha)^* \end{aligned}$$

schreiben, wenn β einen Ausdruck sein sollte, der alle Wörter über $\{a, b\}$ beschreibt, deren Länge durch 4 teilbar ist.

a) Sei $\Sigma = \{0, 1, 2, \ldots, 9, \oplus, \ominus, \textbf{.}, {}^\circ, C\}$. Konstruieren Sie einen erweiterten regulären Ausdruck α über Σ, der genau die *gültigen Temperaturangaben mit zwei Nachkommastellen in Grad Celsius* beschreibt ($0^\circ C$ kann hier mit beiden Vorzeichen versehen sein). Eine solche Temperaturangabe ist gültig, wenn sie

- genau zwei Nachkommastellen besitzt (nach dem Dezimalpunkt **.**);
- keine überflüssigen führenden Nullen im ganzzahligen Anteil aufweist (Gegenbeispiele sind $013.88^\circ C$ bzw. $-00123.45^\circ C$);
- den Minimalwert von $-273.15^\circ C$ nicht unterschreitet und
- mit dem Zusatz $^\circ C$ endet.

Dabei sollen die Symbole $\oplus$ bzw. $\ominus$ anstelle der üblichen Zeichen $+$ und $-$ verwendet werden, um Verwechslungen mit dem Auswahloperator $+$ zu vermeiden: Gültige Wörter sind dann beispielsweise $1234.56^\circ C$, $\oplus 333.99^\circ C$ und $\ominus 199.91^\circ C$. **(2 Punkte)**

b) Seien $\Sigma = \{a, b\}$ und $L_1 = \{\sigma\sigma u \mid \sigma \in \Sigma \text{ und } u \in \Sigma^*\}$ und $L_2 = \{\tau w \tau \mid \tau \in \Sigma \text{ und } w \in \Sigma^*\}$. Geben Sie reguläre Ausdrücke $\alpha_1, \alpha_2, \alpha$ mit $L(\alpha_1) = L_1$, $L(\alpha_2) = L_2$ und $L(\alpha) = L_1 \cap L_2$ an.

(2 Punkte)

c) Nach dem Standard ISO 8601 wird ein Datum in der Form `JJJJ-MM-TT` notiert. Beispielsweise wird der Geburtstag Alan Turings, der 23. Juni 1912, durch `1912-06-23` repräsentiert. Konstruieren Sie einen erweiterten regulären Ausdruck über dem Alphabet $\{0, 1, \ldots, 9, -\}$, der alle gültigen Daten des Jahres 2018 beschreibt. **(2 Punkte)**

Lösung:

a) Ein erweiterter regulärer Ausdruck, der die Sprache erzeugt, ist

$$\alpha = \Big(\underbrace{(\overbrace{\oplus?\beta}^{\text{pos.}} + \overbrace{\ominus\gamma}^{\text{neg.}})}_{\text{bel. Nachkommastellen}}.[0-9]^2 + \underbrace{\overbrace{\ominus\delta}^{\text{neg.}}}_{\text{eingeschr. Nachkommastellen}}\Big)°C$$

mit den Teilausdrücken β, γ und δ, die im Folgenden definiert sind. Sie erzeugen jeweils Repräsentationen bestimmter Zahlen im Dezimalsystem ohne führende Nullen.

- $\boldsymbol{\beta} = \mathbf{0} + [\mathbf{1}-\mathbf{9}][\mathbf{0}-\mathbf{9}]^*$
  Alle natürlichen Zahlen.
- $\boldsymbol{\gamma} = \overbrace{[\mathbf{0}-\mathbf{9}]}^{\text{einstellig}} + \overbrace{[\mathbf{1}-\mathbf{9}][\mathbf{0}-\mathbf{9}]}^{\text{zweistellig}} + \overbrace{\mathbf{1}[\mathbf{0}-\mathbf{9}][\mathbf{0}-\mathbf{9}] + \mathbf{2}[\mathbf{0}-\mathbf{6}][\mathbf{0}-\mathbf{9}] + \mathbf{27}[\mathbf{0}-\mathbf{2}]}^{\text{dreistellig}}$
  Alle Zahlen von 0 bis 272, die mit negativem Vorzeichen und beliebigen Nachkommastellen auftreten können $(0, 1, \ldots, 272)$.
- $\boldsymbol{\delta} = \mathbf{273.}(\mathbf{0}[\mathbf{0}-\mathbf{9}] + \mathbf{1}[\mathbf{0}-\mathbf{5}])$
  Alle erlaubten Werte bei ganzzahligem Anteil -273, mit den erlaubten Nachkommastellen $(00, \ldots, 15)$.

b) Die Sprachen L_1 und L_2 enthalten genau die Wörter über dem Alphabet $\{a, b\}$, deren erste beiden Zeichen übereinstimmen bzw. deren erstes Zeichen mit dem letzten übereinstimmt. Sie werden durch die Ausdrücke $\boldsymbol{\alpha_1} = (\boldsymbol{aa}+\boldsymbol{bb})(\boldsymbol{a}+\boldsymbol{b})^* \equiv \boldsymbol{aa}(\boldsymbol{a}+\boldsymbol{b})^*+\boldsymbol{bb}(\boldsymbol{a}+\boldsymbol{b})^*$ und $\boldsymbol{\alpha_2} = \boldsymbol{a}(\boldsymbol{a}+\boldsymbol{b})^*\boldsymbol{a}+\boldsymbol{b}(\boldsymbol{a}+\boldsymbol{b})^*\boldsymbol{b}$ erzeugt. Die Schnittmenge $L = L_1 \cap L_2$ enthält demnach alle Wörter, deren erste beide Zeichen übereinstimmen und deren erstes Zeichen zudem mit dem letzten übereinstimmt. Erzeugt wird diese Sprache durch den Ausdruck

$$\boldsymbol{\alpha} = \underbrace{\boldsymbol{aa} + \boldsymbol{bb}}_{\textbf{genau 2 Zeichen}} + \underbrace{\boldsymbol{aa}(\boldsymbol{a}+\boldsymbol{b})^*\boldsymbol{a} + \boldsymbol{bb}(\boldsymbol{a}+\boldsymbol{b})^*\boldsymbol{b}}_{\textbf{mindestens 3 Zeichen}}.$$

c) Je nach Monat gibt es 28, 30 oder 31 Tage.

Der folgende Ausdruck ermöglicht die Konkatenation jeder Monatsnummer mit der passenden Anzahl von Tagen.

$$\mathbf{2018} - \big(\mathbf{02} - \boldsymbol{\alpha_{28}} + (\mathbf{04}+\mathbf{06}+\mathbf{09}+\mathbf{11}) - \boldsymbol{\alpha_{30}} + (\mathbf{01}+\mathbf{03}+\mathbf{05}+\mathbf{07}+\mathbf{08}+\mathbf{10}+\mathbf{12}) - \boldsymbol{\alpha_{31}}\big),$$

Dabei sind die Teilausdrücke definiert als

$$\begin{aligned}
\alpha_{28} &= 0[1-9] + 1[0-9] + 2[0-8], \\
\alpha_{30} &= 0[1-9] + 1[0-9] + 2[0-9] + 30 \text{ und} \\
\alpha_{31} &= 0[1-9] + 1[0-9] + 2[0-9] + 3(0+1).
\end{aligned}$$

Eine alternative Lösung ist nachfolgend angegeben. Die Ausdrücke β_{28}, β_{30} und β_{31} beschreiben jeweils die Monate, die *mindestens* die entsprechende Anzahl von Tagen haben.

$$\begin{aligned} \beta_{28} &= 0[1-9] + 1[0-2] \\ \beta_{30} &= 01 + 0[3-9] + 1[0-2] \\ \beta_{31} &= 0(1+3+5+7+8) + 10 + 12 \end{aligned}$$

Aus diesen Teilausdrücken setzt sich, im Sinne einer Fallunterscheidung (vermöge des Auswahloperators), der folgende Ausdruck zusammen, der alle gültigen Daten beschreibt.

$$\mathbf{2018} - \Big(\boldsymbol{\beta_{28}} - (\mathbf{0}[\mathbf{1}-\mathbf{9}] + \mathbf{1}[\mathbf{0}-\mathbf{9}] + \mathbf{2}[\mathbf{0}-\mathbf{8}]) + \boldsymbol{\beta_{30}} - (\mathbf{29} + \mathbf{30}) + \boldsymbol{\beta_{31}} - \mathbf{31}\Big)$$