Prof. T. Nipkow
K. Kappelmann, J. Rädle, L. Stevens **ABGABE 18.05.2020, 23:59 (Σ 5 P.)** Technische Universität München
Lehrstuhl für Logik und Verifikation

Einführung in die Theoretische Informatik

Sommersemester 2020 – Übungsblatt Lösungsskizze 3

AUFGABE 3.1. (*HTML 1:0 RegEx*) 1,5 Punkte

Glaubt man diesem berühmten Stack Overflow-Post, so ist HTML keine reguläre Sprache. So berühmt der Post auch sein mag, gibt er jedoch keinen expliziten Beweis für diese Behauptung. Wir bleiben skeptisch, bis ein Beweis vorhanden ist: Zeigen Sie, dass HTML in der Tat nicht mit einem regulären Ausdruck – wie in der Vorlesung definiert – geparsed werden kann. Sie dürfen die Aussage für eine vereinfachte Variante $\mathsf{HTML}^{\texttt{<b>}}$ über dem Alphabet $\Sigma := \{<, >, /, \mathbf{b}\}$ zeigen, die ausschließlich aus balancierten <**b**>...</**b**> Elementen besteht. Balanciert bedeutet dabei, dass

(a) jeder geöffnete <**b**> Tag auch wieder mit </**b**> geschlossen wird und
(b) ein **b**-Element immer erst geöffnet werden muss, bevor es geschlossen werden kann.

Hier ein paar Beispiele:

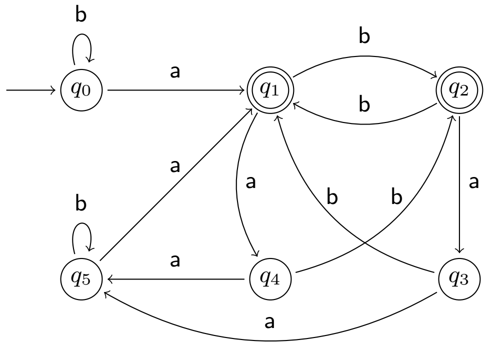
$$<\mathbf{b}></\mathbf{b}> \in \mathsf{HTML}^{<\mathbf{b}>}$$
$$<\mathbf{b}></\mathbf{b}><\mathbf{b}></\mathbf{b}> \in \mathsf{HTML}^{<\mathbf{b}>}$$
$$<\mathbf{b}><\mathbf{b}></\mathbf{b}><\mathbf{b}></\mathbf{b}></\mathbf{b}> \in \mathsf{HTML}^{<\mathbf{b}>}$$
$$<\mathbf{b}></\mathbf{b}></\mathbf{b}><\mathbf{b}> \notin \mathsf{HTML}^{<\mathbf{b}>}$$
$$<\mathbf{b}></\mathbf{b}><\mathbf{b}> \notin \mathsf{HTML}^{<\mathbf{b}>}$$
$$<\mathbf{b}/> \notin \mathsf{HTML}^{<\mathbf{b}>}$$

Lösungsskizze

- Angenommen $\mathsf{HTML}^{<\mathbf{b}>}$ wäre regulär.
- Dann können wir das Pumping-Lemma anwenden.
- Sei $n \in \mathbb{N}$ eine Pumping-Lemma-Zahl.
- Dann gilt $z := (<\mathbf{b}>)^n(</\mathbf{b}>)^n \in \mathsf{HTML}^{<\mathbf{b}>}$ und $|z| \geq n$.
- Es gibt also für z eine Zerlegung $z = uvw$ mit $v \neq \varepsilon$ und $|uv| \leq n$, sodass $uv^iw \in \mathsf{HTML}^{<\mathbf{b}>}$ für alle $i \in \mathbb{N}$.
- Durch unsere Wahl von z folgt, dass uv keine schließende </**b**> Tags enthält.
- Sei $i \in \mathbb{N}$ die Anzahl der <**b**> Tags in v. Wir können v zerlegen in $v = a(<\mathbf{b}>)^i b$ mit $a \in \{\mathbf{b}>, >, \varepsilon\}$ und $b \in \{<\mathbf{b}, <, \varepsilon\}$.
  (a) Falls $i > 0$, enthält uv^2w mehr <**b**> als </**b**> Tags und somit $uv^2w \notin \mathsf{HTML}^{<\mathbf{b}>}$. Widerspruch.
  (b) Falls $i = 0$, betrachte $v^3 = a(ba)^2b$. Wir haben $ba \in \{<\mathbf{bb}>, <\mathbf{b}>, <\mathbf{b}, <>, <, \mathbf{b}>, >\}$. Der Fall $ba = \varepsilon$ ist ausgeschlossen, da $v \neq \varepsilon$.
    (i) Im Fall $ba = <\mathbf{b}>$ enthält uv^3w erneut mehr <**b**> als </**b**> Tags und somit $uv^3w \notin \mathsf{HTML}^{<\mathbf{b}>}$. Widerspruch.
    (ii) In allen anderen Fällen folgt $uv^3w \notin \mathsf{HTML}^{<\mathbf{b}>}$, da uv^3w die invalide Zeichenkette $(ba)^2$ enthält. Widerspruch.
- Alle Fälle führen also zu einen Widerspruch. Das heißt unsere ursprüngliche Annahme ist falsch und $\mathsf{HTML}^{<\mathbf{b}>}$ nicht regulär.

AUFGABE 3.2. 1 Punkt

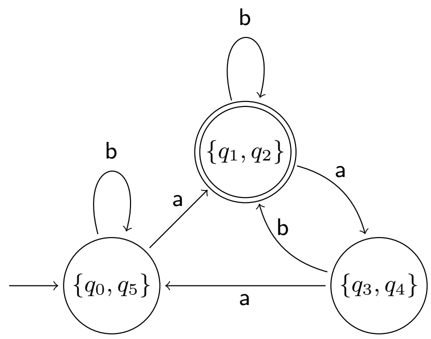
Minimieren Sie den folgenden DFA.



Lösungsskizze
Tabelle:

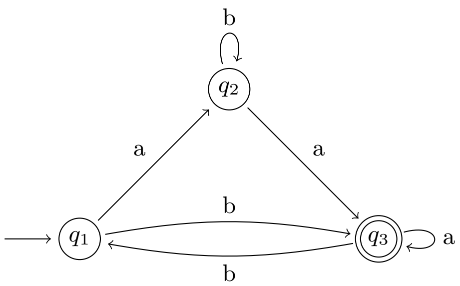
	q_0	q_1	q_2	q_3	q_4	q_5
q_0	–	–	–	–	–	–
q_1	ε	–	–	–	–	–
q_2	ε	=	–	–	–	–
q_3	b	ε	ε	–	–	–
q_4	b	ε	ε	=	–	–
q_5	=	ε	ε	b	b	–

Minimierter DFA:



AUFGABE 3.3. 1 Punkt

Gegeben sei folgender Automat $M = (\{q_1, q_2, q_3\}, \{a, b\}, \delta, q_1, \{q_3\})$:



Berechnen Sie mit dem Gauß-Verfahren und Ardens Lemma einen regulären Ausdruck α mit $L(\alpha) = L(M)$.

Lösungsskizze
Gleichungssystem:

$$X_1 \equiv aX_2 \mid bX_3 \tag{1}$$
$$X_2 \equiv aX_3 \mid bX_2 \tag{2}$$
$$X_3 \equiv aX_3 \mid bX_1 \mid \boldsymbol{\varepsilon} \tag{3}$$

Gleichung (2) nach X_2 auflösen und in (1) einsetzen:

$$X_2 \equiv b^*aX_3 \tag{4}$$
$$X_1 \equiv ab^*aX_3 \mid bX_3 \equiv (ab^*a \mid b)X_3 \tag{5}$$

Gleichung (5) in (3) einsetzen und auflösen:

$$X_3 \equiv aX_3 \mid b(ab^*a \mid b)X_3 \mid \boldsymbol{\varepsilon} \tag{6}$$
$$\equiv (a \mid b(ab^*a \mid b))^* \tag{7}$$
$$\equiv (a \mid bb \mid bab^*a)^* \tag{8}$$

Einsetzen in (5):

$$X_1 \equiv \alpha \equiv (ab^*a \mid b)(a \mid bb \mid bab^*a)^* \tag{9}$$

AUFGABE 3.4. (*Rekursive Prozedur*) 1,5 Punkte

Geben Sie eine rekursive Prozedur $insert(x, r)$ an, die für einen gegebenen regulären Ausdruck r einen neuen regulären Ausdruck berechnet, sodass gilt $\mathsf{L}(insert(x, r)) = \{uxv \mid \exists w \in \mathsf{L}(r). w = uv\}$. Die Sprache $\mathsf{L}(insert(x, r))$

soll also alle Wörter enthalten, die man durch das Einfügen eines Zeichens x in ein Wort aus der Sprache $\mathsf{L}(r)$ erhalten kann. Für Ihre Definition sollten Sie das folgende Gerüst verwenden:

- $\mathit{insert}(x, \boldsymbol{\emptyset}) =$
- $\mathit{insert}(x, \mathsf{a}) =$
- $\mathit{insert}(x, \boldsymbol{\varepsilon}) =$
- $\mathit{insert}(x, \alpha\beta) =$
- $\mathit{insert}(x, \alpha \mid \beta) =$
- $\mathit{insert}(x, \alpha^*) =$

Beweisen Sie mittels struktureller Induktion, dass Ihre Definition korrekt ist, wobei Sie den Fall $r = \alpha\beta$ weglassen dürfen.

Lösungsskizze

Konstruktion

- $\mathit{insert}(x, \boldsymbol{\emptyset}) = \boldsymbol{\emptyset}$
- $\mathit{insert}(x, \boldsymbol{\varepsilon}) = \mathsf{x}$
- $\mathit{insert}(x, \mathsf{a}) = \mathsf{xa} \mid \mathsf{ax}$
- $\mathit{insert}(x, \alpha\beta) = \alpha\, \mathit{insert}(x, \beta) \mid \mathit{insert}(x, \alpha)\, \beta$
- $\mathit{insert}(x, \alpha \mid \beta) = \mathit{insert}(x, \alpha) \mid \mathit{insert}(x, \beta)$
- $\mathit{insert}(x, \alpha^*) = \mathsf{x} \mid \alpha^*\, \mathit{insert}(x, \alpha)\alpha^*$

Korrektheit Wir zeigen $\mathsf{L}(\mathit{insert}(x, r)) = \{uxv \mid \exists w \in \mathsf{L}(r). w = uv\}$ mittels struktureller Induktion.

Fall $r = \boldsymbol{\emptyset}$.
Es gilt $\mathsf{L}(\mathit{insert}(x, \boldsymbol{\emptyset})) = \mathsf{L}(\boldsymbol{\emptyset}) = \emptyset = \{uxv \mid \exists w \in \emptyset.\ w = uv\} = \{uxv \mid \exists w \in \mathsf{L}(\boldsymbol{\emptyset}).\ w = uv\}$.

Fall $r = \boldsymbol{\varepsilon}$.
Es gilt $\mathsf{L}(\mathit{insert}(x, \boldsymbol{\varepsilon})) = \mathsf{L}(\mathsf{x}) = \{uxv \mid w = \varepsilon\} = \{uxv \mid \exists w \in \mathsf{L}(\boldsymbol{\varepsilon}).\ w = uv\}$.

Fall $r = \mathsf{a}$.
Es gilt $\mathsf{L}(\mathit{insert}(x, \mathsf{a})) = \mathsf{L}(\mathsf{xa} \mid \mathsf{ax}) = \mathsf{L}(\mathsf{xa}) \cup \mathsf{L}(\mathsf{ax}) = \{uxv \mid a = uv\} = \{uxv \mid \exists w \in \mathsf{L}(\mathsf{a}).\ w = uv\}$.

Fall $r = \alpha^*$.
Wir erhalten die Induktionshypothese $\mathsf{L}(\mathit{insert}(x, \alpha)) = \{uxv \mid \exists w \in \mathsf{L}(\alpha).\ w = uv\}$. Es gilt

$$
\begin{aligned}
\mathsf{L}(\mathit{insert}(x, \alpha^*)) &= \mathsf{L}(\mathsf{x} \mid \alpha^*\, \mathit{insert}(x, r)\, \alpha^*) = \{x\} \cup \mathsf{L}(\alpha)^*\, \mathsf{L}(\mathit{insert}(x, \alpha))\, \mathsf{L}(\alpha)^* \\
&= \{x\} \cup \mathsf{L}(\alpha)^* \{uxv \mid \exists w \in \mathsf{L}(\alpha).\ w = uv\} \mathsf{L}(\alpha)^* && \text{folgt aus I.H.} \\
&= \{x\} \cup \{uxv \mid \exists w \in \mathsf{L}(\alpha)^+.\ w = uv\} && \text{Beweis folgt unten} \\
&= \{uxv \mid \exists w \in \{\varepsilon\}.\ w = uv\} \cup \{uxv \mid \exists w \in \mathsf{L}(\alpha)^+.\ w = uv\} \\
&= \{uxv \mid \exists w \in \{\varepsilon\} \cup \mathsf{L}(\alpha)^+.\ w = uv\} \\
&= \{uxv \mid \exists w \in \mathsf{L}(\alpha)^*.\ w = uv\}
\end{aligned}
$$

Für die Gültigkeit der obigen Gleichungskette müssen wir noch Aussage

$$\mathsf{L}(\alpha)^* \{uxv \mid \exists w \in \mathsf{L}(\alpha).\ w = uv\} \mathsf{L}(\alpha)^* = \{uxv \mid \exists w \in \mathsf{L}(\alpha)^+.\ w = uv\}$$

rechtfertigen. Wir beweisen die Gleichung, indem wir Mengeninklusion in beide Richtungen zeigen.

Inklusion $\subseteq$
Sei q in $\mathsf{L}(\alpha)^* \{uxv \mid \exists w \in \mathsf{L}(\alpha).\ w = uv\} \mathsf{L}(\alpha)^*$. Dann gibt es $a_1, a_2 \in \mathsf{L}(\alpha)^*$ und $uv = w \in \mathsf{L}(\alpha)$, sodass $q = a_1 u' x v' a_2$. Daraus folgt $a_1 u' v' a_2 = a_1 w a_2 \in \mathsf{L}(\alpha)^+$, womit wir auch $q = a_1 u' x v' a_2 \in \{uxv \mid \exists w \in \mathsf{L}(\alpha)^+.\ w = uv\}$ haben, indem wir $u = a_1 u'$, $v = v' a_2$ und $w = uv \in \mathsf{L}(\alpha)^+$ wählen.

Inklusion $\supseteq$
Sei q in $\{uxv \mid \exists w \in \mathsf{L}(\alpha)^+.\ w = uv\}$. Dann gibt es $u'v' = w' \in \mathsf{L}(\alpha)^+$ mit $q = u' x v'$. Laut Vorlesung gilt $\alpha^+ \equiv \alpha\alpha^* \equiv \alpha\alpha^*\alpha^* \equiv \alpha^*\alpha\alpha^*$. Mit $w' \in \mathsf{L}(\alpha)^+$ gibt es $a_1, a_2 \in \mathsf{L}(\alpha)^*$ und $w \in \mathsf{L}(\alpha)$, sodass $w' = a_1 w a_2$. Indem wir $w \in \mathsf{L}(\alpha)$ geeignet in $w = uv$ zerlegen, erhalten wir sowohl $uxv \in \{uxv \mid \exists w \in \mathsf{L}(\alpha).\ w = uv\}$ als auch $w' = a_1 w a_2 = a_1 u v a_2 = u'v'$. Schlussendlich haben wir $q = u' x v' = a_1 u x v a_2 \in \mathsf{L}(\alpha)^* \{uxv \mid \exists w \in \mathsf{L}(\alpha).\ w = uv\} \mathsf{L}(\alpha)^*$.

Fall $r = \alpha \mid \beta$.
Für $\gamma \in \{\alpha, \beta\}$ erhalten wir die Induktionshypothesen $\mathsf{L}(\mathit{insert}(x, \gamma)) = \{uxv \mid \exists w \in \mathsf{L}(\gamma).\ w = uv\}$. Es gilt

$$
\begin{aligned}
\mathsf{L}(\mathit{insert}(x, \alpha \mid \beta)) &= \mathsf{L}(\mathit{insert}(x, \alpha) \mid \mathit{insert}(x, \beta)) = \mathsf{L}(\mathit{insert}(x, \alpha)) \cup \mathsf{L}(\mathit{insert}(x, \beta)) \\
&= \{uxv \mid \exists w \in \mathsf{L}(\alpha).\ w = uv\} \cup \{uxv \mid \exists w \in \mathsf{L}(\beta).\ w = uv\} \\
&= \{uxv \mid \exists w \in \mathsf{L}(\alpha) \cup \mathsf{L}(\beta).\ w = uv\} \\
&= \{uxv \mid \exists w \in \mathsf{L}(\alpha \mid \beta).\ w = uv\}.
\end{aligned}
$$

Die zweite Gleichung folg aus den Induktionshypothesen.

Fall $r = \alpha\beta$. Die Induktionshypothesen sind wie im vorherigen Fall.

$$\begin{aligned}
\mathsf{L}(\mathit{insert}(x, \alpha\beta)) &= \mathsf{L}(\mathit{insert}(x, \alpha)\,\beta \mid \alpha\,\mathit{insert}(x, \beta)) \\
&= \mathsf{L}(\mathit{insert}(x, \alpha))\mathsf{L}(\beta) \cup \mathsf{L}(\alpha)\mathsf{L}(\mathit{insert}(x, \beta)) \\
&= \{uxv \mid \exists w \in \mathsf{L}(\alpha).\ w = uv\}\mathsf{L}(\beta) \cup \mathsf{L}(\alpha)\{uxv \mid \exists w \in \mathsf{L}(\beta).\ w = uv\} && \text{folgt durch I.H.} \\
&= \{uxv \mid \exists w \in \mathsf{L}(\alpha)\mathsf{L}(\beta).\ w = uv\} && \text{Ähnlich zu Fall } r = \alpha^* \\
&= \{uxv \mid \exists w \in \mathsf{L}(\alpha\beta).\ w = uv\}
\end{aligned}$$

> Parsing HTML with regex summons tainted souls into the realm of the living.
>
> — bobince